9-2017

Song Rhythm Development in Zebra Finches

Julia Hyland Bruno
The Graduate Center, City University of New York

How does access to this work benefit you? Let us know!
Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Behavioral Neurobiology Commons

Recommended Citation
https://academicworks.cuny.edu/gc_etds/2354

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
SONG RHYTHM DEVELOPMENT IN ZEBRA FINCHES

by

JULIA HYLAND BRUNO

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2017
Song Rhythm Development in Zebra Finches

By

Julia Hyland Bruno

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date

Ofer Tchernichovski

Chair of Examining Committee

Date

Richard Bodnar

Executive Officer

Supervisory Committee:

Christopher B. Braun

Mark E. Hauber

Lucas C. Parra

Sarah M. N. Woolley

THE CITY UNIVERSITY OF NEW YORK
This dissertation investigates song-rhythm learning in songbirds. Songbirds have been studied extensively in mechanistic investigations into the sensorimotor underpinnings of the cultural transmission of learned vocalizations. While several studies identified forebrain song-system neurons that generate rhythmic song patterns, we know little about how song rhythms are learned. The first part of the dissertation describes methods for detecting and analyzing birdsong rhythm patterns, and demonstrates their utility for identifying the role of song rhythms in social interactions. Results suggest that rhythm plasticity in zebra finch song may provide a potential vehicle for communication. Controlled song-learning experiments further found that developing zebra finches more readily incorporated a new song element when the tutored rhythm was unchanged, suggesting that a rhythmic framework is established during song learning. An updated schema of the song imitation process is proposed which situates sequence learning within a rhythmic framework. Finally, the role of striatal dopamine in song-mediated social cohesion in zebra finches was identified. Taken together, the dissertation’s findings lay a foundation for future explorations of rhythm in vocal learning and communication.
In memory of

My surrogate grandmother, Stanja Lowe – “Drink-your-tea!”
My first biology teacher, Leigh Van Valen – “Stomp your feet, crack your tail; 6.6 on the Richter Scale!”
My recently-departed maternal uncle, Jake Jacobsen – “Take a liking to a viking.”
ACKNOWLEDGMENTS

This work is the outcome of a series of chance circumstances which led me to Ofer Tchernichovski and his zebra finches, and subsequently to want to become an ethologist. It is with deep gratitude that I acknowledge Ofer’s mentorship over the long haul of the PhD. Looking back, I am grateful in particular for his uniquely creative spirit, his encouragement of difficult projects, and his care for the communal environment in which science is conducted.

In that vein, I can’t really imagine my graduate student time without my lab-mates, who became dear colleagues and friends: Jon Benichov, Olga Fehér, Dina Lipkind, Iva Ljubičić, Primož Ravbar, Tina Roeske, and Kirill Tokarev. I will not soon forget our many adventures. These people have also helped me in crucial and concrete ways. Tina’s experiments on song development in a social arena yielded the co-singing birds whose songs I presented in my first scientific talk and which appear in Chapter 2. I caught my love of birdsong raster visualizations from Dina, who taught me experimental and analytical techniques that I rely on throughout the dissertation. And Kirill was the lead author of the PET imaging study that makes up Chapter 4 (Tokarev et al., 2017). I thank him for giving me permission to include it in my dissertation, and for making the collaboration so enjoyable.

Outside of the lab, I wish to thank my teachers, colleagues, and students connected to the Hunter College Psychology and City College Biology Departments, from whom I have learned much that I will carry forward. I am very proud to have been a part of CUNY.

Throughout my PhD I have also been continually grateful for the intellectual generosity of my “peers” in the international birdsong-science community and beyond. The following individuals/interactions stand out: the late Allison Doupe, who showed patient interest in my first SfN poster and who was an unwitting role model; Arik Kershenbaum, who was so welcoming at the “Analyzing animal vocal sequences” Investigative Workshop at the University of Tennessee’s National Institute for Mathematical and Biological Synthesis in October 2013; Andries Ter Maat, Lisa Trost, Hannes Sagunsky, and Markus
Abels of the Manfred Gahr group in the Max Planck Institute for Ornithology, who hosted Jon and me for a week at Konrad Lorenz’ Seewiesen while they painstakingly taught us how to build zebra finch backpack-microphones; and Jesse Prinz, who provided space and company for speculation in the CUNY Graduate Center Committee for Interdisciplinary Science Studies’ 2015-2016 seminar on animal minds.

For their diverse expertise, time, critical eyes, and very helpful suggestions I thank my dissertation committee members: Sarah Woolley, Lucas Parra, Mark Hauber, and Chris Braun.

Finally, big thanks are due to my friends and family, and especially Adam Weg, my favorite travel companion, for his ardent and multifarious support.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 Zebra finch song development</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2 Timescales of zebra finch song rhythm</td>
<td>2</td>
</tr>
<tr>
<td>1.1.3 Singing and behavioral state</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Overview of the dissertation</td>
<td>5</td>
</tr>
<tr>
<td>CHAPTER 2 REGULARITIES IN ZEBRA FINCH SONG BEYOND THE REPEATED MOTIF</td>
<td>6</td>
</tr>
<tr>
<td>2.1 Abstract</td>
<td>6</td>
</tr>
<tr>
<td>2.2 Introduction</td>
<td>7</td>
</tr>
<tr>
<td>2.3 Method and results</td>
<td>9</td>
</tr>
<tr>
<td>2.3.1 Subjects</td>
<td>9</td>
</tr>
<tr>
<td>2.3.2 Analysis of bout syntax</td>
<td>9</td>
</tr>
<tr>
<td>2.3.3 Temporal regularities beyond bout syntax</td>
<td>11</td>
</tr>
<tr>
<td>2.3.4 Diversity in bout structure repertoires across birds</td>
<td>13</td>
</tr>
<tr>
<td>2.3.5 Evidence for song plasticity during singing interactions</td>
<td>16</td>
</tr>
<tr>
<td>2.4 Discussion</td>
<td>19</td>
</tr>
<tr>
<td>CHAPTER 3 ZEBRA FINCHES LEARN VOCAL SEQUENCES WITHIN A RHYTHMIC FRAMEWORK</td>
<td>22</td>
</tr>
<tr>
<td>3.1 Abstract</td>
<td>22</td>
</tr>
<tr>
<td>3.2 Introduction</td>
<td>23</td>
</tr>
<tr>
<td>3.3 Method</td>
<td>27</td>
</tr>
<tr>
<td>3.3.1 Subjects</td>
<td>27</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

| Table 1. | Results of statistical tests to address the differences in body movement in zebra finch males and females in different conditions: in silence or during conspecific song playbacks. | 83 |
| Table 2. | Results of statistical tests to address the differences in head movement in zebra finch males and females in different conditions: in silence or during conspecific song playbacks. | 84 |
| Table 3. | Results of statistical tests to address the differences in tolerance to air puffs in zebra finch males in different conditions: in silence or during conspecific song playbacks after saline injections, or same after injection of dopamine receptor antagonist L-741,626. | 88 |
| Table 4. | Results of statistical tests to address the differences in tolerance to air puffs in mated zebra finch females in different conditions: in silence and during playbacks of songs of unfamiliar males or their mates. | 90 |
LIST OF FIGURES

Figure 1. Asymmetries in sequences of motif connection types. 11
Figure 2. Timing plasticity in song bouts. 12
Figure 3. Variability in bout repertoire is associated with motif connection type. 13
Figure 4. Evidence for a functional distinction between motif connectors and Introductory notes. 15
Figure 5. Song flexibility at the bout structure level during singing interactions. 17
Figure 6. Experimental paradigm for manipulating target song rhythm. 28
Figure 7. Rhythm of tutored song affects imitation of combinatorial structure. 30
Figure 8. Example songs at baseline and crystallization. 31
Figure 9. Linear and circular representations of the target model song rhythm, illustrating the relationship between syllable inter-onset-intervals and measures of song cycle and phase. 32
Figure 10. Hypothetical relative rhythm imitation at different tempi. 34
Figure 11. Evidence for relative rhythm learning. 35
Figure 12. The altered-rhythm target song model was associated with learning impairments in sequence, rhythm, or both. 37
Figure 13. Within-bird comparison of baseline vs. crystalized tempo stability. 38
Figure 14. Raster view of song development (ISO ex.). 40
Figure 15. Developmental dynamics of song tempo and rhythm (ISO ex.). 42
Figure 16. Integration of new song elements using birds’ own established rhythms. 44
Figure 17. Sequential integration affects song rhythm. 46
Figure 18. Raster view of the developmental transition from the source to the target song (ISO ex.). 47
Figure 19. Dynamic coupling between source and target song cycles (ISO ex.). 49
Figure 20. Song tempo development – individual trajectories. 51
Figure 21. Convergence of source and target song cycles during sequence learning. 52
Figure 22. Raster view of song development (NON-ISO ex.).
Figure 23. Raster view of the developmental transition from the source to the target song (NON-ISO ex.).
Figure 24. Developmental dynamics of song tempo and rhythm (NON-ISO ex.).
Figure 25. Acquisition of isochronous rhythm in a NON-ISO bird.
Figure 26. Acquisition of isochronous rhythm in a NON-ISO bird – with an improvised song element.
Figure 27. Phase diversity, false starts, and period-tripling in a NON-ISO bird.
Figure 28. Song tempo variability development – individual trajectories.
Figure 29. Lasting arrhythmia in NON-ISO birds with developmental discontinuities in song tempo.
Figure 30. Phenomenological categorization of the observed effects of rhythm on song learning.
Figure 31. Delayed PET of dopamine neurotransmission in response to song stimuli.
Figure 32. Song reinforcement assay.
Figure 33. Dopamine neurotransmission in response to song stimuli in unmated males and females.
Figure 34. Song reinforcement in unmated males and females.
Figure 35. Blockage of D2 receptor binding activity by L-741,626.
Figure 36. Song reinforcement after dopamine receptor blockage.
Figure 37. Song reinforcement in mated females.
Figure 38. Dopamine neurotransmission in response to song stimuli in mated females.
1.1 Background

1.1.1 Zebra finch song development

Birdsong is culturally transmitted, typically from father to son (in the majority of species, where females do not sing; cf. Odom et al., 2014), through a slow copying process (Tchernichovski et al., 2001) accompanied by gene expression (Clayton, 1997; White, 2010), hormonal (Marler et al., 1988, Korsia & Bottjer, 1991; Bottjer & Hewer, 1992), and neuronal (Aronov et al., 2008; Ölveczky et al., 2011) changes. For the zebra finch, an opportunistic as opposed to seasonal breeder (Zann, 1996), song development occurs just once, coincident with sexual maturation. A juvenile male zebra finch enters the sensitive period for vocal learning toward the end of his first month, when he begins to attune to and memorize the sounds of the birds around him, produce unstructured babbling, and gradually modify his own vocalizations, using auditory feedback, to match an internal representation (*template*; Konishi, 1965) of the mature song target. By three months, he has acquired the single, idiosyncratic song which he will sing for the rest of his life (which could last ten years or more in captivity), and vocal exploration is thought to cease. Although the basic motif of his final song will likely resemble his tutor's, imitation is never exact, and experimental manipulations have revealed the importance of social factors in shaping the outcome of song development. Thus, although juvenile zebra finches raised in social isolation will copy sounds heard only from recordings, live tutors are preferred (Derégnaucourt et al., 2013); brothers may inhibit each other's learning, leading to divergent imitation within a group (Tchernichovski & Nottebohm, 1998); and juveniles may also copy from one another in addition to or instead of an adult (Derégnaucourt & Gahr, 2013).

Irrespective of copying fidelity, however, the song motif, or basic repeating unit (Immelmann, 1969), of a mature zebra finch exhibits extraordinary temporal precision (on the order of milliseconds).
from rendition to rendition (Chi & Margoliash, 2001). Zebra finch song, therefore, is said to be stereotyped, and song development is often treated as a process of skill perfection, of eliminating variability through practice (e.g., Sakata & Vehrencamp, 2012). This conception, though it has been tremendously useful for investigating the neural mechanisms of sensorimotor learning in the songbird brain, is only true to the extent that it disregards a) timescales beyond the repeated motif; b) behavioral states and social contexts; and c) actual trajectories of development, which are not governed solely by a drive to imitate (Tchernichovski & Marcus, 2014). A focus on skill perfection is not necessarily well-suited for understanding the processes underlying behavioral development, which often follow unexpected and indirect paths (Tchernichovski et al., 2001; Lipkind et al., 2013).

1.1.2 Timescales of zebra finch song rhythm

Compared with birds that have large song repertoires, individual zebra finches may seem to sing the same unique but simple song, the same way every time. For song complexity or plasticity, researchers tend to look to other species such as the thrush nightingale (Rothenberg et al., 2014) or California thrasher (Sasahara et al., 2012). An individual zebra finch song motif typically lasts about half a second and consists of a fixed sequence of 3-7 acoustically-distinct sound segments (syllables) separated by brief gaps of silence. However, zebra finches do not emit isolated motifs when they sing: bouts of song are prefaced by an accelerating series of “introductory notes” (Rajan & Doupe, 2013) and contain a variable number of motif repetitions, punctuated by various call elements and silent intervals analogous to musical rests. The number of introductory notes, motif repetitions, and calls are all variable, and give the song its audible rhythmic structure. Price (1979) recognized that call elements and silent intervals between motifs create a variable structure at the level of the song bout, as did Morris (1954), who likened zebra finch singing to “the turning of a squeaky handle” (ibid, p. 284). However, how the bout-structure variability of zebra finch song performances may be expressed in different contexts has not been explored. In addition, it is unknown how these rhythms develop, or if, how, or when they may be learned.
Even in his most stereotyped bouts of song, the zebra finch is a highly rhythmic singer. His repeating motif provides a strong beat (with interesting timing variations), and the motif itself contains complex temporal as well as acoustic structure (this rhythm can be further reduced to an isochronous beat at smaller timescales; Norton & Scharff, 2016). Developmental perspectives have revealed intriguing temporal dynamics accompanying song acquisition (Tchernichovski et al., 2001; Fehér et al., 2009; Saar & Mitra, 2008; Glaze & Troyer, 2013; Sasahara et al., 2015), but rhythm learning itself has not previously been investigated.

1.1.3 Singing and behavioral state

Although a young bird does not need to interact with an adult tutor in order to accomplish the sensorimotor feat of learning to sing, zebra finches are naturally gregarious songbirds, and the natural context of their song development – and singing – is the social group. Zebra finches in the Australian grasslands spend most of their lives in large flocks, surrounded by hundreds of other birds (Zann, 1996). Unlike territorial species, zebra finches do not sing to ward off rival males (ibid). Males use song to court females and maintain pair-bonds (zebra finches mate for life; ibid), but they are often seen singing in larger social settings, which early observers speculated could facilitate flock cohesion (Immelmann, 1968, 1969; Hall, 1962; Sossinka & Böhner, 1980), a hypothesis which has never been tested.

Morris (1954) originally described a continuum of high- and low-intensity courtship song and “stationary song” according to the vigor of the bird’s accompanying rhythmic pivoting movements and their involvement of the whole body or just the head. Sossinka & Böhner (1980) dichotomized female-directed and undirected song types on the basis of acoustic differences. They recorded the songs of domesticated and wild-caught zebra finches singing alone and with an unfamiliar female and found that the courtship context was associated with more introductory notes, shorter motif duration, and more motifs per strophe (bout). The bimodal distributions seen in these song features were interpreted to reflect the importance of joint stimulus- and motivational control in “releasing” singing behavior of variable intensity. In other words, the behavioral differences were taken as an indication of mechanistic differences, the prediction being that undirected song is gated by a lower threshold. In support of this
hypothesis, Walters et al. (1991) found that blocking testosterone aromatization inhibits directed but not undirected song. Hessler and Doupe (1999) further reported dampened female-directed singing-related neural activity in the anterior forebrain pathway (AFP; a basal ganglia-thalamo-“cortical” loop), which generates variability required for developmental vocal learning (Aronov et al., 2008), and Jarvis et al. (1998) found that AFP immediate early gene expression is also lower during directed vs. undirected singing.

Thus, while it was originally associated with immobility, undirected singing came to be regarded as a form of vocal practice, in contrast to the stereotyped ‘performance’ context of directed song. It is unknown how these context-dependent aspects of singing might vary with different audiences, including other males. However, early investigators (Sossinka & Böhner, 1980; Bischof et al. 1981) assumed that zebra finch courtship song would show gradations in intensity as a function of the reaction of the courted female, and a positive feedback relationship was observed between male singing and female bowing (Garson et al., 1980). Given the apparent communicative function of back and forth calling exchanges among zebra finches of both sexes (Elie et al., 2010; Benichov et al., 2016), bidirectional responsiveness between males and females during male courtship song would not be surprising. If a male zebra finch is capable of responding to feedback from social interactions and can make real-time adjustments to his “stereotyped” song, this biologically-significant behavior may be more rightly viewed as an engagement than a performance.
1.2 Overview of the dissertation

The unifying motivation for this thesis is an interest in understanding how vocal learning occurs within the context of the complex social environment into which the young bird is born. Of course, I have not exhausted this question in my PhD. Each of the three chapters of this dissertation tackles an aspect of this guiding question, with a focus on temporal organization, a feature of song learning latent in the literature, but as yet underexplored.

Chapter 2 examines the structure and function of rhythms in adult zebra finch song beyond the level of the imitated motif. As these timescales of song rhythm have not been characterized before, I present methods for detecting and analyzing song bout structure, and demonstrate their utility for identifying the role of song rhythms in social interactions not restricted to the courtship setting. Evidence suggests that rhythmic plasticity in zebra finch song may provide a potential vehicle for communication. A complementary set of results is presented in Chapter 4, which reports experimental findings that listening to the songs of other males is behaviorally-reinforcing for male zebra finches, and also increases striatal dopaminergic activity in a PET imaging assay with awake, behaving birds.

The core chapter (Chapter 3) focuses on song rhythm from the perspective of the developing bird. In controlled song learning experiments designed to decouple rhythm and sequence learning, learning outcomes and developmental analyses indicate that birds must learn their songs within a rhythmic framework. The implications of these findings for existing models of song learning and song timing control in the songbird brain are discussed in the context of a new phenomenological schema.
CHAPTER 2
REGULARITIES IN ZEBRA FINCH SONG BEYOND THE REPEATED MOTIF

2.1 Abstract

The proliferation of birdsong research into the neural mechanisms of vocal learning is indebted to the remarkable stereotypy of the zebra finch’s song motif. Motifs are composed of several copied syllables, which birds learn to produce in a fixed order. But at a higher level of organization – the bout – zebra finch song is no longer stereotyped. Song bouts include several repetitions of the motif, which are often linked by a variable number of short “connector” vocalizations. Here we show that combinatorial analysis yields an incomplete description of this bout-level structure. In contrast, studying birdsong as a time-varying analog signal can reveal patterns of flexibility in the rhythmic organization of bouts. Visualizing song bouts as sorted raster plots of acoustic features shows that motifs are strung together via two distinct categories of connections: tight or loose. Loose connections allow considerable timing variation across renditions. Even among co-tutored birds that acquired the same motif, we observe strong individual variability in the temporal plasticity of song bouts. Our investigation suggests that vocal flexibility could potentially allow individuals to express a variety of behavioral states through their songs, even in species that sing only a single stereotyped motif.

---

1 A version of this chapter will appear as part of a Birdsong Learning special issue of Behavioural Processes (Hyland Bruno & Tchernichovski, in press).
2.2 Introduction

Birdsong is among the most thoroughly studied vocal communication systems in non-human animals. However, relatively little is known about how the complex acoustic structure of birdsong might relate to specific social functions (cf. Todt & Naguib, 2000). Zebra finches are the predominant model system for studying vocal learning, due in part to the remarkable stereotypy of their songs. The male zebra finch typically produces renditions of his song 'motif' in bouts that begin with a series of introductory notes, followed by several renditions of the same motif. An individual song motif includes several 'syllables,' each with a distinct and complex acoustic structure. Both the syllable types and their combinatorial order within the motif are highly stereotyped (Immelmann, 1969; Scharff & Nottebohm, 1991). Therefore, after hearing an adult zebra finch producing even a single song bout, one already knows much about his song structure, which remains stable over years. In other words, although a zebra finch often sings a lot (hundreds of motifs every day), his songs are nearly identical; it is as if the bird is always saying the same thing, again and again. To the extent that this picture is true, zebra finch song would seem unlikely to carry much dynamic information—for instance, to express the behavioral state of the bird, or to direct different intentions to other birds. This is puzzling, however, given that zebra finches are highly social, and appear to be communicating vocally much of the time (Elie & Theunissen, 2015).

In fact, we know that the zebra finch song motif is not entirely stereotyped, and does carry some information about behavioral state, as is observed in female-directed versus undirected singing. When a male zebra finch courts a female, he produces a dance, and his song motifs change a bit—for example, motifs are sung slightly faster (Sossinka & Böhner, 1980), and syllable acoustic structure becomes slightly more precise (Kao & Brainard, 2006). Further, female zebra finches can perceive these differences, and they typically prefer the female-directed version of the song (Woolley & Doupe, 2008). Still, the acoustic differences between female-directed and undirected songs are very small, and some of these can be explained by (involuntary) changes in brain temperature (Aronov & Fee, 2012). In sum, only a very restricted degree of plasticity has been found to exist in the mature zebra finch’s song motif.
However, zebra finch song can only be considered stereotyped at the level of the motif, i.e., in repeating short sequences of syllables. The next level of song organization – the song bout – is not stereotyped: First, the number of motif renditions varies, typically ranging from two to ten motifs per bout. Second, song bouts often vary in the manner in which motifs are strung together. Figure 1a presents five song bouts produced by one bird. As shown, the motifs are linked via a variable number of short vocalizations (Price, 1979); we call these “connectors.” One approach for investigating the structure of song bouts would be to treat the song bout as a sequence of symbols and study its combinatorial structure. For example, the bout segments shown in Fig. 1a can be described as sequences of motifs (‘M’) and connectors (‘i’), e.g., M-i-M-ii-M. With this approach, one can estimate the transition probabilities between motifs and connectors within the bout. First-order Markov Models have been shown to be useful for describing birdsong structure, although there is some debate about the appropriateness of such models (Kershenbaum et al., 2014). In canaries, for example, a recent study uncovered long-term temporal dependencies between phrases unexplainable by simple ‘bigram’ Markov models (Markowitz et al., 2013). An entirely different approach is to investigate the song bout as an analog time-varying signal, focusing not on ‘syntax’ but on temporal regularities such as rhythms. Here, we compare the two approaches, considering transition probabilities between symbols and the continuous time structure of song bouts side by side.

One may wonder whether it is even appropriate to treat the combinatorial (symbolic) and time-varying (analog) structure of birdsong as distinct levels of song organization. In linguistics, combinatorial structure (e.g., grammar, based on symbolic units) and temporal regularities (e.g., prosody, measured from the acoustic speech signal) are studied by different scientific communities, as they represent plausibly distinct features of language. In birdsong, however, we do not know if, or to what extent, song syntax and rhythms are independent levels of organization (Mol et al., in press). If these two levels are coupled, as our preliminary findings will suggest, then variability in song sequences could mirror adjustments of rhythms, serving, for example, to coordinate singing behaviors across individual birds. More generally, it might be essential to combine sequential analysis with investigation of temporal...
regularities in order to understand the communicative function of the song bout. In the following we focus on presenting methods for exploring zebra finch bout structure, and raise hypotheses for future research.
2.3 Method and results

2.3.1 Subjects

Birds in this study were adults, between 134 and 271 days post-hatch (dph). All experiments were approved by the Hunter College Institutional Animal Care and Use Committee.

2.3.2 Analysis of bout syntax

In order to first estimate the combinatorial structure of song bouts, we examined the transitions between song motifs within bouts. Figure 1 presents such transitions in 289 song bouts produced by one adult zebra finch (270 dph) over the course of a day. In this bird’s repertoire, motifs (‘M’) were in some cases linked directly to each other (M-M), but more often, motifs were strung together via short connector notes (‘i’). Example sequences of motifs and connectors are shown in Fig. 1a. The bird produced different types of connections, based on the number of connectors: M-M, M-i-M, M-ii-M, etc. These connection types occur in sequences, e.g., M-i-M-ii-M, which we regard as a transition from the M-i-M type to the M-ii-M type. Figures 1b and c present the expected and observed pairwise transition probabilities between connection types (within three-motif sequences). We see two effects: first, the frequencies of pairwise transitions between bout subunits showed a much narrower distribution than expected from the frequency of connection types (Fig. 1d, $\chi^2 (36, n = 407) = 209.62, p < .001$). That is, there was an apparent rule-governed or ‘syntactic’ regularity in the transitions. Second, the pairwise transitions showed strong asymmetry: M-i-M, the most common subunit, was often followed by M-ii-M (M-i-M-ii-M frequency, 20.6%). However, the opposite transition, from M-ii-M to M-i-M, was rare (5.2%). This order-dependent asymmetry in the transition probabilities between shorter and longer connection-type subunits was stable across days and was statistically significant for this bird ($\chi^2 (20, n = 290) = 144.20, p < .001$).
2.3.3 Temporal regularities beyond bout syntax

Why were song motifs of this one bird often linked together via a single connector and then two connectors, but rarely the other way around? Analysis at the level of the time course of the song bout provides additional information for elucidating the observed syntactic regularities. Figure 2 shows raster plots of the song bouts in this bird’s full repertoire (as recorded on days 270 & 271 post-hatch), sorted by bout type and duration. A conservative 300-millisecond silence criterion was used to segment bouts (Okubo et al., 2015; Norton & Scharff, 2016); this threshold encompassed over 80% of all silent gaps under 10 seconds. Pseudo-colors in Figs. 2a-c represent millisecond-resolution pitch changes within different syllables. Motifs appear as white and orange bands in the raster; connectors were lower in pitch,
Figure 2. Timing plasticity in song bouts. a, Three representations of a song bout (same bird as in Fig. 1): top, symbolic string of motifs ('M') and connectors ('i'); middle, sonogram; bottom, millisecond-resolution pitch vector such as used to construct raster plots of large song samples. The color scheme used to represent 100-1000 Hz pitch values in a-c is indicated by the color bar. b, Sorted raster plot of all the songs produced by this bird over two days, showing one bout per row. Bouts are aligned to the beginning of the first motif and sorted to reveal connection type and temporal jitter within motif/connector sequences. c, Closer view of the area in b (blue box) showing greater relative plasticity when motifs were joined by two vs. one connector. d, Histograms of the intervals separating motifs and subsequent song elements. Intervals in cases where motifs were linked by zero, one, two, or three connectors are shown in separate panels, with the y-axis indicating relative frequency with respect to all the motif transitions in the bird’s repertoire.

and stand out as red strips between motifs. As can be seen, when two motifs were joined via a single connector, the connection was very tight. However, when motifs were linked together by two connectors, the distribution of gaps was much more variable across renditions (Fig. 2c). Figure 2d presents histograms of time intervals between the end of the previous motif and the beginning of the first connector (or first motif element in cases where no connectors were used). As shown, for both M-M and M-i-M, gaps were short and their distributions were narrow (Fig. 2d, top panels). In contrast, when motifs were linked via two or three connectors (M-ii-M, M-iii-M), the time interval between the motif end and the next song element was much more variable (Fig. 2d, bottom panels). We therefore call the single connector ‘tight’ and the double and triple connectors ‘loose.’ Exhaustion could perhaps explain the asymmetries we
observed in bout syntax; i.e., once a bout becomes loose, it is unlikely to become tight again. An alternative hypothesis is that song plasticity can be flexibly controlled.

### 2.3.4 Diversity in bout structure repertoires across birds

Figure 3 presents sorted raster plots of bout repertoires from eight additional birds that were raised together with a single tutor (as in Tchernichovski & Nottebohm, 1998). At the end of song development (134-145 dph), these birds had acquired similar song motifs (with modest variations), but they varied considerably in how motifs were strung together. Comparing bout repertoires across birds

![Figure 3](image_url)

**Figure 3. Variability in bout repertoire is associated with motif connection type.** Individual panels present sorted bout raster plots showing all songs produced in a single day for eight adult birds (ages 134-145 dph) that were raised together with the same tutor. Bouts are aligned to the onset of the first motif, and colors indicate pitch. Bird identity is included in the upper right corner of each panel. Birds on the left (p5, p4, p1, and p7) exhibited little variability in the timing of their motifs, as seen in the vertical columns in the raster plots, whereas birds on the right showed greater variability.
(Fig. 3, different panels), raster plots appear ‘step-like’ in some birds (e.g., in bird p5) and gradated in others (e.g. in bird p2). At the combinatorial level of motif transitions, some birds (p5, p4, p1) produced primarily single connectors \((M-i-M)\), others (p7, p2, p9) primarily two connectors \((M-ii-M)\), and a third group produced both.

Single connectors were tight in all birds that produced them. Motifs strung together with two connectors tended to be loose – but to a variable extent across birds. For example, p7 and p9 both used two connectors exclusively, but for p7 motif connections were invariably tight (visible in the vertical columns in the raster), whereas p9 produced two acoustically-distinct types of double connectors, one loose and the other tight (with the first of the two connectors in the tight variant marked by a pitch downsweep).

Connectors were often acoustically distinct from introductory notes (which are used to introduce song bouts). Further, in several birds, the acoustic (spectral) structure of connectors varied across bout types. We therefore distinguish between connector type (spectral), and connection type (the number of connectors between motifs). Pitch differences among connectors can be seen for example within the bouts of birds p9 and p6. Indeed, most birds possessed repertoires of multiple acoustically-distinct connectors (Fig. 4, red clusters). In these eight birds, 13 out of 24 total connector types formed non-overlapping clusters with introductory notes (shown in blue, Fig. 4), and likewise 10 of 22 introductory note types were unique to the starts of bouts.
Figure 4. Evidence for a functional distinction between motif connectors and introductory notes. Individual panels present scatter plots showing duration vs. goodness of pitch (an estimate of harmonic pitch periodicity [Tchernichovski et al., 2000]) for both introductory notes (blue) and connectors (red), for each of the birds shown in Fig. 3. The opacity for all scatter plot data points is set at 5% to indicate density. Individual panels are labeled by bird identity, follow the same order as Fig. 3.

These findings provide support for the integrity of the song bout as a biologically meaningful unit, as opposed to an arbitrary (observer-imposed) segmentation of singing behavior. We see moreover that there are two means of loosening the stereotypy of connections between motifs: by having a bout vocabulary consisting of more than one type of motif connection (via variable numbers of connectors and/or variable connector types), and through temporal jitter in the timing of the connectors themselves. As a result, some birds sing with less predictability than others. Possessing a variety of options for constructing song bouts could allow a bird to adapt his singing behavior based on context or feedback from a social partner. While it is beyond the scope of this study to test the hypothesis that loose motif connections may serve a social function, using an example will outline methods for exploring song bout plasticity in social scenarios.
2.3.5 Evidence for song plasticity during singing interactions

We selected one bird with rigid song bouts, and another bird with high temporal variability of song bouts (p5 and p6 from the group shown in Fig. 3). We placed them together for several days and monitored their singing interactions. Figure 5 presents a snapshot of the pair’s singing interactions during two consecutive days. These birds, which had shared a tutor, possessed similar motifs but dissimilar bout structure repertoires. Bird p5’s singing style consisted of stereotyped bouts of 1-3 motifs, predominantly separated by a single connector with only tight connections between motifs (left raster plot and motif/connector histogram, Fig. 5a). Bird p6, by contrast, produced song bouts with greater temporal variability and including both tight and loose motif connections (Fig. 5a, right). Both birds sang a lot (p5: 3,928 motifs in two days, p6: 2,790 motifs) but relatively rarely (~11% of all bouts) at the same time. When their songs did overlap, however, there was a clear asymmetry in the tendencies of the two birds to co-sing (see examples in Fig. 5b).

Figure 5c shows a raster view of all the overlapping songs. For each bird, we sorted the bouts anchored to the onset of the partner’s bouts (Figs. 5c, left and right panels respectively). As indicated by the horizontal black demarcation lines, bird p5, who performed exclusively tight motif connections (blue), tended to initiate singing, while bird p6, who performed both tight and loose motif connection (red), tended to join in. Note that the slope of the raster (marked 1, 2 & 3 in Fig. 5c, left) represents the likelihood of joining. We can see that singing likelihood increased steeply with the bout onset of p5 and declined rapidly after p5’s first motif. This indicates that p6 tended to join (co-sing) during p5’s first motif. Irregularities in the timing of p6’s subsequent motifs (Fig. 5c, left) also suggest preferred bout patterns, perhaps related to the predictable time structure in the songs of p5 (Fig. 5c, right). Specifically, during the tightest co-singing episodes (section 2 in Fig. 5c, left), we can see that many of p6’s second motifs within a bout begin near the offset of p5’s second motif, or the onset of p5’s third motif; the variation in gap duration visible between p6’s first and second motifs is due to the use of variable connection types. These motif latency dynamics were observed on each of the two days (Fig. 5d), and histograms of relative bout onsets only (Fig. 5e) clearly distinguish p6 as the ‘answering bird’ (positive latency) relative to p5 (negative latency). While songs produced simultaneously between these two birds were not always
Figure 5. Song flexibility at the bout structure level during singing interactions. a, Solo song bout raster plots (top) and motif/connector interval histograms (bottom; see Fig. 2d) for p5 (left) and p6 (right), two birds raised in the same social group (Fig. 3). Intervals between motif offsets and subsequent connector onsets are plotted separately for motifs linked by a single vs. more than one connector (M-i-M vs. M-ii-M, M-iii-M, etc.). b-d, Co-singing data for birds p5 (blue) and p6 (red). b, Symbolic raster plot of five co-singing examples showing alignment of introductory notes, motifs and connectors. c, Symbolic raster plots (motifs only) of all overlapping songs occurring over the course of two consecutive days, with the data sorted by the bout onset lag of either p6 (left) or p5 (right). Horizontal black lines in both panels demarcate bouts initiated by one bird versus the other. Slopes of the numbered sections of the left raster are as follows: 1) -.11, 2) -.39, and 3) -.09, indicating that the probability of p6 to join in was greatest during the first motif of p5. d, Probability density estimates of each bird’s motif onsets relative to the start of the partner’s bout, plotted separately for the two days shown in c. e, Probability density estimates of the relative timing of each bird’s bout onsets across both days.
synchronized, coordinated singing (e.g., Fig. 5b) would only be possible, given each bird’s potential for bout plasticity (Fig. 5a), if p6 were to join in with p5, rather than the other way around. Given that this pair of birds developed their songs together, it is an intriguing question whether the establishment of bout structure repertoires might be related to patterns of social interaction.
2.4 Discussion

The song bout is the first timescale at which zebra finch song is no longer stereotyped, making it a logical place to investigate potential communicational aspects of singing behavior. Here, we have taken a first step toward characterizing the variability that exists in the structure and timing of zebra finch song performances.

Behaviorally, we have long known from Price (1979) and others that the song bout is an important level of song organization (for example, the courtship dance is organized at the level of the song bout [Williams, 2001; Ullrich et al., 2016]). Studies of the neuronal coding of birdsong, however, have focused mostly on the hierarchical coding of syllable timing and acoustic structure in the premotor nuclei HVC and RA (e.g., Yu & Margoliash, 1996; Hahnloser et al., 2002). Still, there is some evidence for bout-level neural coding in the premotor song system. Williams and Vicario (1993) observed “superbursting” upstream of HVC in the thalamic nucleus Uva at the termination of song bouts. Chi and Margoliash (2001) reported changes in intra-syllable RA burst intervals as a function of motif position within the bout. And recently, Okubo et al. (2015) found a subset of HVC projection neurons that are active exclusively at bout onsets. How does the songbird brain gate which bout variant is produced from a repertoire of possible types, and what determines the timing of motifs (tight vs. loose connections)?

Rajan and Doupe (2013) showed that song-system neurons (including HVC interneurons and Area X projections) encode the serial position of introductory notes, which accelerate and converge on a stereotyped acoustic state signaling readiness to begin singing. They also found that the number of “introductory notes” is correlated with the time elapsed since the end of the previous song motif, when considering intervals up to one second – i.e., including what we refer to here as “connectors.” Mechanistically, how does initiating a song bout differ from initiating a motif within a bout? Given our finding that motif connectors are often acoustically distinct from introductory notes, it could make sense to repeat the study by Rajan and Doupe (ibid), but distinguishing between different introductory note and connector types. Perhaps the acoustic differences we observed could be explained the brain’s varying
“readiness” to produce sequences of stereotyped learned vocalizations (motifs) as a function of whether the bird is already in a singing “state.”

We hope that future studies will soon reveal which brain areas regulate song bout plasticity. Sossinka & Böhner (1980) first dichotomized female-directed and undirected song on behavioral grounds, showing that the courtship context was associated with more introductory notes, shorter motif duration, more motifs per bout, and greater sequential stereotypy. The observed context-dependent shifts in these song features were interpreted to reflect the importance of joint stimulus- and motivational control in “releasing” singing behavior of variable intensity. This phenomenon has since been substantiated across multiple mechanistic levels (Walters et al., 1991; Hessler & Doupe, 1999; Jarvis et al., 1998). Kao & Brainard (2006) further demonstrated that anterior forebrain pathway (AFP) lesions abolish acoustic feature variability found to accompany undirected song, although there was no effect found on the structural differences in the two types of song. It may be that temporal jitter in motif connections is also mediated by the AFP, while choice of bout ‘vocabulary’ (i.e., connector type) is controlled elsewhere. Since zebra finches can acquire multiple motif variants by adding or deleting syllables (Sturdy et al., 1999; rare examples of truncated motifs are visible in Figs. 2b and 2c above), neuro-ethological investigations of bout structure variability may want to consider both connectors and motif variants.

At the functional level, an in-depth characterization of singing behavior in social context, including tracking production of tight and loose connections, is needed in order to figure out the possible role of bout structure and plasticity in expressing behavioral state and coordinating behaviors. In many passerine species, song serves at least two social functions: courtship display, and male-male territorial defense (Kroodsma, 2005). However, there is no evidence that this is true for zebra finches, which are not territorial (Zann, 1996). Although attacks and fighting are not unusual in captive settings, singing behavior does not appear to be associated with agonistic interactions in the lab or in the wild (Evans, 1970; Caryl, 1975; Immelmann, 1969). The primary biological function of the zebra finch’s song is thought to be sexual, playing an important role in both courtship and pair-bond maintenance (zebra finches pair for life [Zann, ibid]). Somewhat mysteriously, then, wild male zebra finches spend proportionally more time
singing when no females are present (Dunn & Zann, 1996). This “undirected” song is by no means all solitary, and singing in the presence of other males was once speculated to facilitate flock cohesion (Immelmann, 1968, 1969; Hall, 1962; Sossinka & Böhner, 1980). While Immelmann reported male zebra finches “singing in duos and trios” as early as 1968, we know of only one paper since then that mentioned male-male co-singing (Anisimov et al., 2014). The tight and loose song motif connections we observed could be involved in regulating such social functions.

In order to assess the potential for communication via bout structure variability, it will be important to test the influence of social partner behavior on song performance. Heinig et al. (2014) found that male Bengalese finch song syntax varied in the presence of different females. Might the analog methods we propose reveal a similar capacity for performance adaptation in the zebra finch on the level of timing? New wireless recording technology that enables source separation of the vocalizations of individual birds in social groups (Ter Maat et al., 2014; Anisimov et al., 2014) should be used in future systematic study of song plasticity. Such technology could also be used to probe whether bout structure is meaningful to the listener, for example by monitoring the timing of female calls during a potential male’s song over the course of pair bonding. Another interesting angle would be to investigate the influence of the tutor’s bout structure on song learning in the developing juvenile, given prior evidence that pupils copy “chunks” of syllables corresponding to occasional production breaks in the tutor’s song (Williams & Staples, 1992).

Finally, are any of the bout structure features we reported here learned? What could be the origin of the bout structure diversity we observed? We showed that birds that learn the same song together as a group can nevertheless develop divergent bout structure repertoires. What developmental processes generate this diversity and what is the time course of bout structure ontogeny? Are bout structures learned from a tutor during the sensitive period for song development? Here we presented only snapshots of adult song repertoires; we do not know how age or experience might affect singing behavior in adulthood. Might learning how to sing be separate from learning what to sing? Longitudinal data will be key to elucidating the dynamic information that zebra finch song may carry despite its stereotypy.
3.1 Abstract

Birdsong is composed of ordered sequences of song syllables. Chains of neuronal activity in the songbird brain’s song nuclei generate each song syllable, and additional chaining mechanisms are thought to link syllables into song phrases. If song syllables are strung together via a sequence generator mechanism, then successful sequencing should not depend on relative duration. We found, however, that birds are much more successful in adding a new syllable type to a song if the duration of the new syllable can fit into a pre-existing rhythm. We trained juvenile zebra finches to change their songs by incorporating a new syllable, which either fit or deviated slightly from the prior rhythm. This deviation had striking effects on the learning outcome: first, birds were less likely to fully accomplish the learning task. Second, examination of song development shows that new syllables were often incorporated into pre-existing rhythmic slots. These findings suggest that the song system is first of all a rhythm generator, and that vocal sequences are learned in the context of preexisting rhythm templates.
3.2 Introduction

The acquired song motif of the adult male zebra finch is a brief (~1-s) rhythmic vocal pattern consisting of a fixed sequence of acoustically-distinct vocal sounds called song-syllables (Zann, 1996). Zebra finch songs are highly stereotyped and syllables are delivered with temporal precision on the order of milliseconds (Chi & Margoliash, 2001). This crystalized stereotypy is all the more remarkable given the complexity of song development (Immelmann, 1969; Tchernichovski et al., 2001). At about one month of age, the young male zebra finch begins to attune to the mature songs in his environment. He encodes an internal representation (template) of adult song (Konishi, 1965), and also begins to produce unstructured vocal babbling, known as subsong. The bird then gradually modifies his vocalizations to match the template, a process requiring auditory feedback. By the end of sexual maturation and the sensitive period for song learning, around 100 days post-hatch, his song has become a high-fidelity imitation of an external model.

Since Nottebohm (Nottebohm et al., 1976; Nottebohm & Arnold, 1976) first identified the forebrain ‘song system’ of discrete nuclei in the songbird brain, birdsong neuroscience has become a successful model system for studying the neural control of motor sequencing and sensorimotor learning. An important advance in understanding the mechanisms of sensorimotor learning in vocal development was the discovery of two descending pathways between the premotor nuclei HVC (used as a proper name) and RA (robust nucleus of the arcopallium). HVC is a sensory-motor nucleus in the neo-striatum, which functions as a primary generator of song patterns. Nucleus RA functions similarly to a mammalian motor cortex. It controls brainstem motoneurons, which drive respiratory and syringeal (bird’s vocal organ) muscles. The first pathway between HVC and RA is called the direct posterior pathway. It is a mono-synaptic connection, which is necessary for producing learned song in the adult bird. The second pathway is called the anterior forebrain pathway (AFP). This complex pathway loops through the basal ganglia and the thalamus before reaching its final target at nucleus RA. The AFP is necessary for song learning (Bottjer et al., 1984) and consolidation (Scharff & Nottebohm, 1991; Aronov et al., 2008).
While early studies posited a direct role for AFP in template-matching (Doya & Sejnowski, 1995), the memory of the tutor’s song likely resides elsewhere (Phan et al., 2006; Hahnloser & Kotowicz, 2010; Roberts et al., 2012). Several studies, however, implicate AFP in generating biased variability in the premotor circuitry (Andalman & Fee, 2009), which produces variable plastic song (Goldberg & Fee, 2011; Ölveczky et al., 2005; Ölveczky et al., 2011) as well as adaptive residual variability in adulthood (Kao et al., 2005; Tumer & Brainard, 2007; Ravbar et al, 2012). AFP-mediated spike-timing dependent plasticity in RA (Sizemore & Perkel, 2011) is a promising candidate mechanism for learning (Farries & Fairhall, 2007; Fiete et al., 2007), and AFP receives performance-related dopaminergic modulation (Gadagkar et al., 2016). Thus AFP appears to be the locus of reinforcement learning, which has been theorized to be a driving force in song development (Troyer & Doupe, 2000; Fee & Goldberg, 2011; Nick, 2015).

A second body of research has explored the neural control of song timing. In a landmark study, Hahnloser et al. (2002) found that individual HVC-RA projecting neurons are clocks, producing a short burst of spikes only once per song motif, always in the same position in song-time. Evidence suggests that these ultra-sparse bursts are produced in a collectively sequential pattern, completely covering the duration of the song motif at millisecond resolution. This effect was further corroborated (Kozhevnikov & Fee, 2007), including in studies using focal cooling of HVC (Long & Fee, 2008; Andalman et al., 2011), which caused near uniform temporal retardation across song. There is some tension between these findings and other results suggesting hierarchical organization of the song system, in which HVC is thought to encode distinct vocal gestures (Vu et al., 1994; Yu & Margoliash, 1996; Amador et al., 2013). It remains to be seen whether the two views are mutually exclusive. In any event, according to the ‘clock’ hypothesis (Fee et al., 2004; Leonardo & Fee, 2005; Glaze & Troyer, 2007; Fee & Scharff, 2010) accurate song time patterns are generated via intrinsic synfire chains of HVC neurons (Jin et al., 2007; Li & Greenside, 2006; cf. Danish et al., 2017) which form a self-contained central pattern generator that provides a feed-forward temporal anchor for the sequential activation of motor commands encoded in RA. Recent studies (e.g., Hamaguchi et al., 2016) challenged the claim that song-timing is restricted to HVC, but evidence for a continuous representation of time in the song-system (either restricted to HVC or not) is overwhelming (Lynch et al., 2016; Picardo et al., 2016). The model of neural chaining in HVC via syn-
firing chains has also been extrapolated to explain the serial production of song syllables (Jin et al., 2007; Jin, 2009; Andalman et al., 2011), and a recent study linked the development of HVC bursting with the emergence of syllables in developing birds (Okubo et al., 2015).

The focus on sequences, whether in investigations of the learning of vocal gestures or of time-coding via sequences of neuronal chains, or even in theories of song error corrections via the basal ganglia (e.g., Andalman & Fee, 2009), is rooted in “the problem of serial order in behavior,” as Karl Lashley put it in 1951 (cited, e.g., in Jin et al., 2007, Troyer & Doupe, 2000). However, according to Lashley, the problem with this problem was that associative chaining was an inadequate explanation for complex sequential behaviors such as speech – and (he speculated) “the song of birds” (Lashley, 1951, p. 113). Lashley’s first, now uncontroversial, point was that the behaviorist concept of reflex chains could not explain context-dependent associations between disparate units in behavioral sequences, e.g., words in a sentence. But he advocated, in turn, for the investigation of integrating schema, such as rhythms (Martin, 1972). The near-perfect sequential and timing stereotypy of zebra finch song has perhaps obscured the need to consider the problem of motor sequencing beyond the serial ordering of units (cf., Glaze & Troyer, 2006; Saar & Mitra, 2008; Glaze & Troyer, 2013; Norton & Scharff, 2016). Meanwhile, other fields dealing with the problem of serially-ordered behavior have long recognized that timing is crucial in the learning of motor skills (Viviani & Terzuolo, 1980; Sakai et al., 2004).

Despite the accumulation of knowledge about mechanisms of birdsong learning, the nature of the internalized song memory ‘template’ representation as well as the mechanisms through which this illusive template guides vocal learning remain mysterious. Is song rhythm learned, or is it an epiphenomenon of the learning of temporal sequences? Here we designed controlled song learning experiments to directly test this question. Taking advantage of the long sensitive period for zebra finch song learning, we used an altered-target training method (Lipkind et al. 2013; Ravbar et al., 2012) to experimentally induce sequence learning with or without the added requirement of learning a new rhythm. During the first month, birds learned to imitate a regularly-timed (isochronous), monosyllabic song sequence (AAAA). Once this first song was acquired, a new song model was introduced that contained an additional syllable
(ABAB). For half of the birds, the duration of the new syllable was shortened slightly, rendering the new target song non-isochronous. From a sequence-learning perspective, the learning task was the same for all birds. If, however, rhythm is learned independently from sequential content, we would expect to see differences in the learning outcomes of the two groups. We tracked the entire vocal development of each bird, to examine how the sequence learning task would be negotiated under conditions in which rhythm is preserved or altered.
3.3 Method

3.3.1 Subjects

A total of 26 male juvenile zebra finches from the Hunter College breeding colony were used in this study. All birds were kept with parents and siblings until 7 days post-hatch (dph), at which point the father was removed to prevent song exposure during the sensitive period for learning. At 31 dph, birds were moved to individual home-cages contained inside sound attenuation chambers, where they remained for the remaining two months of their sensitive period for song development (maximum age recorded: mean, 122.5 dph; range: 106-135 dph). Animals were kept on a 12:12 photo-period schedule and given water, seed, wet food (seed mixed with water and egg powder), and grit ad libitum. Perches, mirrors, and tutor-bird dummies were provided for enrichment. Experiments were conducted with approval from the Hunter College Institutional Animal Care and Use Committee.

3.3.2 Sound recording, playback training and sound analysis

Birds’ vocalizations were recorded continuously throughout the duration of the experiments (Tchernichovski et al., 2004; Tchernichovski et al., 2000). Beginning 35 dph, birds were exposed to playbacks of the source song model (see Experimental Design), with two daily quotas of 15 randomly timed songs (morning and afternoon sessions; playbacks activated at .005 odds/sec). After 9 days, operant keys were introduced to each bird’s cage (Tchernichovski et al., 2000) and playback delivery was switched from passive to self-elicited (key-activated). Playbacks were then reduced to two daily quotas of 10 self-elicited songs (Tchernichovski et al., 1999).

Sound Analysis Pro software (http://soundanalysispro.com; Tchernichovski et al., 2000) was used for sound recording and preliminary analyses. A random ten percent of all sound files recorded each day were selected for automated sound segmentation and acoustic feature extraction using the Sound Analysis Pro batch processing function. Sound Analysis Pro was also used to perform semi-automated clustering to identify song syllables over developmental time. Subsequent data analyses were performed
using Matlab 8 (Mathworks Inc.) custom scripts. A 300-millisecond silence criterion was used to define song bouts.

3.3.3 Experimental design

We manipulated tutor song rhythm using an altered-target training method (Lipkind et al., 2013; Lipkind & Tchemichovski, 2011; Ravbar et al., 2012). First, all birds were exposed to one week of passive playbacks. We then provided them with a key that they could peck in order to trigger playbacks of the same song. The song consisted of regularly-timed (isochronous) repetitions of a single syllable (source song, Fig. 6). We selected 18 out of 26 birds, where a recognizable replica of the source song could be clearly identified in the developing song before day 70. In those birds, playbacks were then switched to the target, a bisyllabic song which had either isochronous (ISO) or non-isochronous (NON-ISO; see Fig. 6) time structure. Birds were randomly assigned to one of either the ISO or NON-ISO groups (9 birds per group).

![Sonograms of song 1 (top) and song 2 (bottom) showing preserved and altered rhythm.](image)

**Figure 6. Experimental paradigm for manipulating target song rhythm.** Pictured are sonograms (time vs. frequency representations) of the monosyllabic source (top) and bisyllabic target (bottom) song models used in Experiment 1. All birds were trained on playbacks of Song 1 (source song) from 35 days post-hatch. As soon as this song was acquired, the song model was switched to Song 2 (target song), which contained an additional syllable and two new transitions as a result (A-B and B-A). Song 2 was either isochronous (like Song 1), or contained a shorter B-syllable which made the rhythm non-isochronous. Blue and red vertical lines indicate alignment between source and target song syllables in the two conditions.
3.4 Results

3.4.1 Song learning outcomes

We trained two groups of 9 birds with an altered-target songs design as in Lipkind et al. (2013). Both groups were presented with the same combinatorial task: first to learn to produce a monosyllabic song (AAAA) and then to alter their song by inserting a new syllable (AAAA source → ABAB target; Fig. 6). In one group, the source and target songs were of the same isochronous rhythm (ISO group, N = 9 birds). That is, the new syllable B was congruent with the existing rhythm. In a second group (NON-ISO, N = 9 birds), the source song (AAAA) was isochronous, but the target song was of slightly non-isochronous rhythm (with the B-A inter-onset-interval [IOI] 44 milliseconds shorter than A-B). In order to assess sequence imitation, we analyzed the proportion of alternating A and B syllable types (ABA & BAB) in the learned song (Fig. 7a). With perfect imitation, the song would be composed strictly of alternating ABA and BAB trigrams.

As shown in Fig. 7b, birds trained on the isochronous ABAB model ended up with high levels of alternating trigrams in their songs (median = 75%), whereas birds in the NON-ISO group were much more heterogeneous (median = 53%, less than the lower quartile for the ISO group). The difference between the two groups was statistically significant (Wilcoxon Rank Sum Test (two-sided), $W_s = 63$, $z = -1.988$, $p = .047$, effect size $r = -.469$). While four NON-ISO birds achieved combinatorial learning outcomes comparable to their isochronous-only counterparts, three NON-ISO birds learned only one of the two new transitions (A-B or B-A), and completed their songs with novel song elements. The remaining two NON-ISO bird failed to incorporate the new syllable into their song. In sum, a minor change in model song rhythm from isochronous to non-isochronous was sufficient to interfere with the sequence-learning task.
Figure 7. Rhythm of tutored song affects imitation of combinatorial structure. a, Schematic demonstrating the procedure used for quantifying song sequence match to the target model song: song performances were transformed into symbolic strings (see Methods), each of which was scanned left to right for 3-syllable sequences matching the ABA and BAB trigrams of the ABAB song model. b, Degree to which birds imitated the target model syllable sequence, quantified as the percent of all 3-syllable sequences consisting of alternating trigrams (ABA or BAB). Individual data points are overlaid on kernel density distributions (violin plots) for each group (isochronous vs. non-isochronous song model).
Samples of the learning outcomes from the two groups are given in Fig. 8. The bird in Fig. 8c, for example, achieved poor sequence imitation both because he generated a novel syllable ('X'), and because he retained many source A-A transitions in his crystalized song.

![Figure 8. Example songs at baseline and crystallization. Shown are example sonograms together with symbolic coding of syllable sequences for the songs of one ISO (a) and two NON-ISO (b, c) birds. Each bird’s imitation of the source song before the target model was switched is displayed above his crystalized song at the end of learning. Age is indicated in days post-hatch for all examples.](image)

We next looked into how birds in the two groups imitated the temporal structure of the new song model. The ISO and NON-ISO target models, while sharing the same sequential structure, differed in both song *tempo* and *rhythm*. As illustrated in Fig. 9, *tempo* is defined by the duration of the song cycle \( \omega \), or period between repeating segments. In the target ABAB song,

\[
\omega = |AB| + |BA|,
\]
where $|AB|$ represents the inter-onset-interval, IOI, between syllables A and B (‘A-B IOI’), and $|BA|$ represents the IOI between B and A. That is, the song cycle is simply the IOI between successive A-syllables. The ISO and NON-ISO song models had cycles of 572 and 528 ms, respectively. As distinct from tempo, song rhythm is defined by the phase angle ($\varphi$) of the onset of the B-syllable, relative to the cycle:

$$\varphi = \frac{|AB|}{\omega}.$$ 

Rhythm is therefore measured as a ratio, as opposed to tempo, which describes the absolute timing of

---

**Figure 9.** Linear and circular representations of the target model song rhythm, illustrating the relationship between syllable inter-onset-intervals and measures of song cycle and phase. a, Schematic of the ABAB target song syllable and gap durations, together with respective syllable IOIs for the isochronous (top, blue B-syllable) and non-isochronous (bottom, red B-syllable) models. b, The two song models now represented as repeating cycles (IOIs between successive A-syllables). The A-B IOIs of the two models have the same duration, but the phase angle of the B-syllable onset differs, because NON-ISO B-syllable is shorter. The NON-ISO song cycle is also shorter than the ISO cycle, as represented in the figure by the relative sizes of the circles.
song elements. The tutored rhythm was characterized by a phase angle of $\pi$ in the ISO group, as compared with $\phi = 3.4 \ (13\pi/12)$ radians in the NON-ISO group.

We considered two alternative hypotheses:

**Hypothesis 1**: Birds simply imitate and combine sequences. That is, they copy the new syllable, including its duration, but they learn each time interval independently. If rhythm is unimportant in the learning of a new sequence (which would be inconsistent with our results so far), then tempo and rhythm imitation error should be uncorrelated (see below).

**Hypothesis 2**: Birds copy the rhythm of the model, including the phase angle of the novel B-syllable onset relative to the song cycle (see Fig. 9). If rhythm is imitated, then the deviation of the bird's A-B IOI from that in the model should scale with the difference in cycle duration between the bird and the model. If this hypothesis is correct, then we expect that some birds should copy the phase angle but not the tempo (as in the diagram in Fig. 10).
Figure 10. Hypothetical relative rhythm imitation at different tempi. Shown are three different hypothetical scenarios for copying the same example ABAB model song rhythm displayed at left. (See Fig. 9 for an explanation of the rhythm schematic.) When rhythm is imitated absolutely (top right), the time intervals in the model are reproduced exactly. Alternatively (middle and bottom right), the same rhythm (phase) may be reproduced at a faster or slower tempo by scaling the two intervals down or up together.

To test between these two hypotheses, we compared tempo and rhythm imitation error in the mature songs of each bird with respect to his target model (ISO or NON-ISO). Tempo error is calculated as simply:

$$\varepsilon_{\text{Tempo}} = \bar{\omega}_{\text{Bird}} - \bar{\omega}_{\text{Model}},$$

where $\bar{\omega}_{\text{Bird}}$ is the bird's median cycle duration. For direct comparison with tempo, we calculate rhythm error based on the absolute timing of the B-syllable (as opposed to the relative phase):

$$\varepsilon_{\text{B\_Timing}} = |AB|_{\text{Bird}} - |AB|_{\text{Model}},$$

where $|AB|_{\text{Bird}}$ is the bird's median A-B IOI.

Figure 11 (a-b) shows the relationship between cycle duration and B-syllable timing across all birds at song crystallization (i.e., between the denominator and the numerator of the B-syllable phase-
Cycle duration (tempo) vs. B-syllable onset timing

Rhythm learning outcomes: B-syllable onset phase angle

Figure 11. Evidence for relative rhythm learning. a-b, Scatter plot showing the relationship between tempo (median cycle duration) and rhythm (median A-B IOI) imitation error for all birds that acquired bisyllabic songs (a, ISO group N = 9 [all birds], b, NON-ISO group N = 5). The ISO and NON-ISO model tempo and rhythm are plotted with solid dots (red and blue, respectively). In a, the blue diagonal line indicates the ISO relative rhythm across the range of tempi shown; the same is plotted with a dashed blue line in b, where the NON-ISO relative rhythm is indicated by the solid red diagonal line. c-d, Crystallized song rhythm, measured as the phase angle of the onset of the new syllable (or syllables, cases where a bird added more than one new syllable to his source song). Polar histograms show pooled data for all birds within each group (c, ISO; d, NON-ISO; n = 1,000 randomly selected song cycles per each of 9 birds per group), with the phases of the isochronous and non-isochronous models drawn in blue and red, respectively.

angle ratio). As shown, ISO birds (Fig. 11a) tended to cluster along the blue diagonal line, indicating that individual birds accurately copied the relative rhythm of the model at varying tempi. In the NON-ISO condition (Fig. 11b), our sample is limited to the five birds that acquired a strictly ABAB song; these birds showed a similar pattern of relative rhythm imitation with varying tempo, with the exception of one bird
that acquired an isochronous relative rhythm rather than his own model’s non-isochronous rhythm. Across all birds, there was a significant positive correlation between tempo and rhythm imitation error (as defined above), $r = .58$, $p = .03$. These results provide support for Hypothesis 2 over Hypothesis 1, according to which the two temporal errors should be uncorrelated. The findings demonstrate that zebra finches are capable of imitating a relative rhythm in a new tempo, and allow us to consider relative phase imitation irrespective of absolute match to the model tempo.

To compare relative rhythm imitation in the ISO and NON-ISO conditions, we computed phase across 1,000 randomly selected cycles in each bird’s mature songs, including any novel syllables generated in addition to or in place of the B-syllable copied from the model. Across pooled songs produced by birds in the ISO group, B-syllable phase angles were tightly distributed near $\pi$ (Fig. 11c). ISO songs had a median +/- median absolute deviation (M.A.D.) phase-angle of $3.17 +/- 0.27$ radians, as compared with $3.44 +/- 0.48$ radians for NON-ISO songs, and the two distributions differed significantly in terms of their variance (Brown-Forsythe test for homogeneity of variance, $F(1, 17998) = 861.21$, $p < .0001$). Indeed, the NON-ISO distribution showed two peaks (Fig. 11d): a primary peak near the B-syllable phase angle of the NON-ISO target model, and secondary peak at $\pi$, suggesting a tendency for NON-ISO birds to incorporate the new song element into the relative rhythm of their original, isochronous song. Only three NON-ISO birds accurately copied both the combinatorial structure and the relative timing of the ABAB target song (Fig. 12).
Figure 12. The altered-rhythm target song model was associated with learning impairments in sequence, rhythm, or both. Scatter plot shows the relationship between B-syllable (or other non-A-syllable) phase angles \((n = 1,000 \text{ song cycles/bird})\) and overall sequence imitation (% ABA or BAB trigrams), within individual birds in both groups. Opacity for all data points is set at 1%. Horizontal black lines represent the individual interquartile ranges of the phase angles. Data from individual birds have been jittered minimally (preserving rank) in the vertical axis for improved visual discrimination. The B-syllable phase angles of the target song models are indicated by the blue and red vertical reference lines at \(x = \pi\) (ISO) and 3.40 radians (NON-ISO).

In addition, several NON-ISO birds also ended up with target songs that were less temporally stable than their original source songs. Within NON-ISO birds, there was a significant difference between cycle median absolute deviation at crystallization and baseline, paired t-test, \(p = .03\). Individual ISO birds showed no such acquired arrhythmia and recovered levels of stability comparable to baseline in their mature target songs (Fig. 13).
**Figure 13. Within-bird comparison of baseline vs. crystalized tempo stability.** Lines compare individual tempo stability in ISO (blue, N = 9) and NON-ISO (red, N = 9) birds at the model switch day vs. after song crystallization (last full day recorded). Tempo stability is measured as song cycle median absolute deviation (M.A.D.) in each bird’s baseline source (AA) and endpoint target (AB) songs (n = all songs recorded at each time point).

To summarize the combinatorial and rhythm outcomes we observed, birds more readily added a new song element when the rhythm of the song model was preserved, and appeared to have difficulty switching to a new rhythm. In addition, birds accurately imitated rhythm but not tempo. Finally, our experimental perturbation appeared to have a lasting effect on tempo stability, as seen in the mature arrhythmia among NON-ISO birds. These song learning outcomes are incompatible with a purely sequential learning model. To investigate how a 44-ms difference in the duration of the target model B-syllable resulted in these divergent song learning outcomes, we examined song changes continuously throughout development.
3.4.2 Analyzing song rhythm development

Figure 14 presents a raster visualization of all the songs produced during the entire learning period by one representative bird from the ISO group, from five days before the song model was switched. (See the Appendix for full developmental raster plots for all birds in the study.) A-syllables are coded in white and B-syllables in blue in this visualization, and aligned song bouts are displayed in chronological order, from top to bottom. Comparing the upper and lower portions of the raster plot, we see that this bird fully adopted the target song. Within the first few days of exposure to the new target model, the bird began to perform occasional B-syllables (in blue) at the starts of song bouts, confirming an “edge effect” as reported in Lipkind et al. (2013). B-syllables did not begin to appear in their mature sequential context (connected to two A’s) until the eleventh day of hearing the new model. Within five days, the transition from the AAAA to the ABAB song was virtually complete.

What happened during this transitional period (Fig. 14, green box)? We can see that the transition to the target was gradual rather than immediate, at least in this example. We wondered, however, whether the target song rhythm might have nevertheless been established immediately, or at least before sequence learning was complete. In the ISO target model, the new B-syllable occupied an existing A-syllable timeslot, and the target song cycle was simply a period-doubled version of the source. If rhythm is indeed operational in song learning, then ISO birds could also reuse existing timeslots in their source songs to incorporate the new syllable – a mechanism unavailable to NON-ISO birds. This process could still result in gradual song learning: by way of analogy, imagine the process of getting into a game of double-dutch jump-rope. If a jumper properly entrains with rhythm of the ropes before leaping, she can land at the right time on her first attempt. However, a novice might need to watch and wait longer than a pro in order to synchronize accurately. Similarly, ABA sequences might be rare initially, but still rhythmically accurate. Alternatively, incorporating the B-syllable at the appropriate rhythm might be more like learning to play darts through target practice. This scenario, in which the gradual acquisition of skill precision actually depends on initial variability (error), would be more in line with a reinforcement-driven model of song learning (Dhawale et al., 2017). Such an account does not predict rhythm learning.
**Figure 14. Raster view of song development (ISO ex.).** Shown is the full song ontogeny of bird R4647 from the ISO group. Each row represents a single song bout rendition, aligned to the onset of the first A-syllable in the bout, with A-syllables plotted in white and B-syllables in blue (song elements that did not cluster well with either A or B are colored gray). Song bouts are presented chronologically, from 5 days prior to the song model switch through to song stabilization, top to bottom. The green box outlines the song transition days, focusing on the first motifs within bouts (see Fig. 18).

 independent of sequence, nor does it predict a difference in learning difficulty between the ISO and NON-ISO conditions.

---

2 A note about the raster plots for bird R4647: due to an idiosyncratic amplitude dip within this bird’s A-syllable, A-syllables were sometimes mis-segmented. In those cases, the acoustics of the syllable were such that the first segment of the syllable was more easily identified by the clustering algorithm than later segments. Because syllable sub-segments could not all be reliably clustered, it appears visually in the raster as if there was variability in the duration of the A-syllable (white bars). This is, however, only a segmentation artifact. It does not impact any of the analyses.
Which process – entrainment or trial-and-error exploration – better describes how birds acquired the target song in our experimental task? We tracked tempo and rhythm changes throughout development by plotting daily distributions of individual song cycle (e.g., Fig. 15a) and B-syllable phase angle (e.g., Fig. 15b). To obtain an invariant metric of tempo development, we measured a double-cycle for birds’ source songs, and compared this ‘cycle’ with the cycle of birds’ emerging target songs. If period-doubling had occurred, then a bird’s source (so defined) and target cycles would form a continuous developmental trajectory. Indeed, continuity of tempo can be seen in the ISO example given in Fig. 15a, where the target song emerges in the same cycle range as the source (notably, this bird sang at a slower tempo than the model throughout development). Prior to disappearing, however, the source song tempo fluctuated several times; these smooth fluctuations show that a steady tempo can nevertheless drift over longer timescales. In juxtaposition (Fig. 15b), the bird maintained a stable imitation of the ISO model rhythm (π), apparently from the first appearance of the target cycle.
Figure 15. Developmental dynamics of song tempo and rhythm (ISO ex.). All data are from ISO bird R4647 (Fig. 14). Developmental trajectories of daily distributions of song cycle (a) and B-syllable phase angle (b). In a, the bird’s source song cycles (‘AA’ timeslot, the IOI between an A-syllable and the A after
next) are plotted in black, with target song cycles (‘AB’ timeslot, the IOI between an A-syllable and the next A following B) in green. Dashed and solid vertical green lines indicate the time points (in days post-switch) at which the bird’s song sequences exceeded 25 and 50% alternating trigrams (ABA or BAB, as in the target song). Opacity is set at 5% for all data points. The model cycle duration and phase are plotted as horizontal blue reference lines.

These data therefore provide anecdotal support for entrainment over trial-and-error rhythm learning. In the next section, we investigate the re-usage of preexisting timeslots in incorporating the new syllable. As a preview, when we measured the phase-angle of the very earliest B-syllables across birds, early rhythm learning did not prove to be very accurate. However, recalculating the phase with reference to the cycle of birds’ source songs revealed that the ISO birds were in fact recycling existing timeslots to learn the new song. Tempo continuity is explored further in section 3.4.4; this turned out to be a more important and general feature of song development than we anticipated.

3.4.3 Timeslot reuse

To investigate early rhythm learning, we examined the timing of the B-syllable just as birds began to adopt the new target song (the first day, for each bird, where the target song sequence reached 5%). We found considerable variability among birds’ median rendition-to-rendition B-syllable phase angles at this very early time point of target cycle formation (Fig. 16a), contrary to the entrainment rhythm-learning hypothesis proposed above. To isolate the timing of the B-syllable from the timing of the closure of the song cycle, we next recalculated phase with respect to three different cycles:

\[ \varphi_B = \frac{|AB|}{\omega_{model}}, \]

where \( \omega_{model} \) represents the cycle of each bird’s target model;

\[ \varphi_C = \frac{|AB|}{\tilde{\omega}_{AA, switchday}}, \]

where \( \tilde{\omega}_{AA, switchday} \) represents the median cycle of the bird’s source song prior to exposure to the ABAB target; and

\[ \varphi_D = \frac{|AB|}{\tilde{\omega}_{AA, current}}, \]

where \( \tilde{\omega}_{AA, current} \) represents the median cycle of the bird’s current source song. We used two different reference frames to relate the phase of the B-syllable to the bird’s source cycle to account for potential
Figure 16. Integration of new song elements using birds’ own established rhythms. a, Polar histograms showing the distribution of the median B-syllable phase angle across birds (ISO, left; NON-ISO, right), at the first appearance of B in context (ABA trigrams at 5%). b, Median B-syllable phase angle calculated using the target model cycle duration as the denominator. c, Median B-syllable phase angle calculated with the bird’s own early (model switch day) source (AA) cycle duration in the denominator. d, Median B-syllable phase angle calculated with the bird’s own current source (AA) cycle duration in the denominator.

drifts in song tempo during development (Fig. 15a). Comparing these three phase projections (Fig. 16b-d), only the ISO birds’ current source song cycle (\(p_d\)) provided a denominator with a strongly unimodal distribution – at \(\pi\) (Fig. 16d, left). This tells us that, among ISO birds, the new B-syllable did appear within a source-song timeslot, but that variability in the newly-formed target cycle made this difficult to discern from phase alone.
Taking the model cycle as denominator ($\varphi_p$) failed to reduce phase variability in the direction of the target rhythm for either group (Fig. 16b), confirming that birds do not copy absolute time intervals from a model. In the NON-ISO group, birds’ own source cycles ($\varphi_c, \varphi_d$; Fig. 16c-d, right) performed better (with a slightly stronger effect found with reference to the switch-day source cycle; Fig. 16c, right); however, these distribution were bimodal. interestingly, NON-ISO birds tended to introduce the B-syllable either at $\pi$ relative to their source cycle (thus attempting to reuse preexisting timeslots, like the ISO birds), or phase-delayed compared to the NON-ISO model rhythm. We can only speculate as to the cause for this spontaneous phase-delay (from the $13\pi/12$ of the NON-ISO model toward $5\pi/4$), but perhaps these results reflect an intrinsic preference for simple rhythms, i.e., small-integer ratios, as has been shown to exist in humans (Jacoby & McDermott, 2017).

Given the diversity of performed phases produced (Fig. 16a) we hypothesized that in some ISO cases the early B-syllable might not match the duration of A. Indeed, ISO bird R4720 provides such an example (Fig. 17). What is most striking about this example, however, is that the duration of the B-syllable was found to depend on its placement in the sequence. We color-coded all acoustically-clustered B-syllables based on whether (blue) or not (pink) they occurred inside a target-song cycle, i.e., in between two A’s. As shown, when the B-syllable began to be incorporated into the song (around 7 days post-switch), it shortened in duration. Moreover, the two B-syllable distributions were distinct at many time points, although they appeared coupled (and both underwent a period of instability when a transient longer-duration cluster was formed between days 10 and 17). Specifically, beyond the 50% target sequence mark, until the song became stereotyped around day 39, the B-syllable was much more variable in duration when it was not embedded within the song cycle. This example demonstrates that sequential context can affect the timing and duration of syllables, thereby providing supporting, if indirect, evidence of a rhythmic framework which constrains the developmental process of song sequencing.
Sequential integration affects song rhythm. Shown are daily distributions of A and B syllable duration over developmental time for bird R4720 from the ISO group. Syllable B is plotted in blue when it occurred between two A’s, otherwise it is pink. The green line indicates ABA trigrams > 50%.

Our finding that ISO B-syllables were accurately timed at their first appearance within the target song sequence shows that rhythm is established early rather than gradually molded. Although birds copied the source model at different tempi, they appear to have perceived a rhythmic equivalence between the source and target model rhythms, and implemented an equivalent rhythmic substitution in their own songs. This picture is in line with an entrainment explanation of rhythm learning, however with an important qualification. Continuing the jump-rope analogy, the variability we observed in cycle-closure in the earliest renditions of the new song sequence means that after successfully ‘jumping in’ birds had yet to learn how to keep jumping in time with the ropes – how not to ‘stumble out.’ By the end of song development, ISO birds had accurately acquired isochronous ABAB songs (Fig. 11c). We next examine song cycle development to see how this was achieved.
3.4.4 Preservation of self-tempo

Of particular interest is the evolution of tempo during the period in which the target song sequence was replacing the source. Figure 18 provides a closer raster view of the songs during the transitional period of the example ISO bird presented in Figs. 14 and 15. Again, A-syllables are plotted in white and B-syllables in blue. In Fig. 18a the song bouts are plotted in chronological order, as in Fig. 14, here focusing on the first three syllables within each bout (beginning with the first A-syllable of the bout;

![Figure 18](image_url)

Figure 18. Raster view of the developmental transition from the source to the target song (ISO ex.). Song bout raster plots (n=816 song renditions) showing the song transition days for bird R4647 (Fig. 14, green box; see also: Fig. 19). a, Aligned song bouts (first 1-second) in chronological order. b. Same data as a, now sorted by the identity of the second syllable in each bout (showing only bouts containing more than one syllable).
occasional unincorporated bout-initiating B-syllables are visible to the left of x = 0). Contained within interleaved white and blue syllable columns are AAA, ABA, and AAB sequences. Sorting this raster by the identity of the second syllable of the bout (Fig. 18b), we can see now that the B-syllable is incorporated in between A-syllables (never in repeating strings). Further, the relatively vertical alignment of the third column (Figs. 18a-b) suggests alignment of the different sequence types and therefore continuity of tempo.

To examine this more carefully, we compared the bird’s source and target song cycles during the transitional period where the two songs coexisted (Fig. 19a). The breakdown of the two distributions on each day of the transition, shown in Fig. 19b, reveals an interesting progression involving adjustments to both cycles. Just before the incorporation of B (day 10 post-switch), the source cycle was distributed fairly tightly around 600 ms (slower than the model tempo). On the next day, when the bird’s imitation of the target first appeared, the source cycle shifted slightly, becoming longer and more diffuse. In subsequent days, the target cycle tightened as the spread of the source distribution increased until day 14, when the two distributions were largely overlapping; the source song then disappeared, and bird shifted to sing the target AB-song exclusively. Comparing the two and bottom distributions (days 10 and 15), we find that the tempo of the bird’s source song was largely preserved in the new target song.
Figure 19. Dynamic coupling between source and target song cycles (ISO ex.). Data are from bird R4647 (see also: Figs. 14, 15 and 18). a, Closer view of the song-cycle data presented in Fig. 15a, days 4 through 25 post-switch (with data point opacity now set at 25%). Pink brackets indicate source (black) and target (green) song cycle data from the same day. b, Distributions of the bird’s source and target song cycle durations on each day during the transition from the source to the target song.
Figure 20 presents average song tempi throughout development for all the birds in both the ISO (Fig. 20a) and NON-ISO (Fig. 20b) groups. As with the raw data cycle trajectories above (Figs. 15a and 19a), birds’ source song cycles are plotted in black, target song cycles in green, and sequence learning timelines are indicated by vertical lines. In order to assess dynamic coupling between the two cycles across birds, we focused on their period of overlap (see Fig. 21a) and calculated the daily difference between each bird’s median source and target cycle durations over time in relation to individual sequence learning timelines (Figs. 21b-c). We found that as birds approached the day on which the target sequence reached 50%, source and target song cycles tended to become more similar in duration.
Figure 20. Song tempo development – individual trajectories. Separate panels show the development of the source (black) and target (green) median song cycles for each bird (ISO, a; NON-ISO, b). Individuals are sorted according to endpoint sequence imitation (proportion alternating trigrams, printed in green in the upper right corner of each panel), top left to bottom right.
**Figure 21.** Convergence of source and target song cycles during sequence learning. **a,** Example of source and target song cycle developmental trajectories from one bird (see **Fig. 20**), with a box outlining the transitory period where the bird produced both songs. **b-c,** Absolute differences between birds’ median source and target song cycles over time, with individual trajectories anchored (day zero) to the day at which target song trigrams (ABA or BAB) reached 50%. For individual birds, the period of song overlap was measured from the first appearance of ABA trigrams to the disappearance of AAA trigrams, or end of development if the source song was never eliminated. Data points from individual birds are plotted in the same color, ranked within each group (see legend) according to endpoint sequence imitation. **b,** ISO birds (n = 9). **c,** NON-ISO birds (n = 5, all that achieved at least 50% target trigrams at least transiently during development).

As shown (**Fig. 21c**), this convergence occurred among NON-ISO as well as ISO birds. In the NON-ISO model, however, the target cycle diverged from the source. How then can we understand these results? And how do they compare with song tempo development among NON-ISO birds that failed to learn the target song sequence (central and bottom three panels in **Fig. 20b**)?

Let us first take a closer look at the best imitator of the NON-ISO condition (**Fig. 20b**, upper left panel). The emergence of a stereotyped ABAB song is clearly visible in the developmental raster view...
shown in Fig. 22. As with the ISO example presented above (Fig. 14), we can also see that the transition between the source and target songs occurred gradually – in this case, close to two weeks.

Figure 22. Raster view of song development (NON-ISO ex.). Shown are song development data for bird R5106 from the NON-ISO group. Presentation is as in the ISO example, above (Fig. 14) except that here the B-syllable is colored red. Aligned song bouts are displayed chronologically from top to bottom. The green box marks the transitional days shown in closer view in Fig. 23, below.

Zooming in on the first syllables of the song bouts produced during the days of the target song’s first appearance (days 12-17 post switch; Figs. 23a), again we see roughly vertical columns of interleaved source and target sequences (compare with the ISO example in Fig. 15). When we sort the
Figure 23. Raster view of the developmental transition from the source to the target song (NON-ISO ex.). Song bout raster plots (n=1022 song renditions) showing the first song transition days for bird R5106 (see Fig. 22, green box). a, Aligned song bouts (first 1-second) in chronological order. b, Same data as a, now sorted by the identity of the second syllable in each bout (showing only bouts containing more than one syllable).

bouts raster data by the identity of the second syllable (Fig. 23b), we can see that the A and B syllables are unequal in duration, as in the NON-ISO model. However, unlike the model, they align relatively vertically in the raster because the silent gap between A and B is longer than that between A and A. This is important because it represents a deviation from the target model rhythm, since the silent gaps of the source and target model songs were all equivalent (Fig. 9). Indeed, the bird’s B-syllables were initially phase-delayed relative to the NON-ISO model (Fig. 24b); the gap was also slightly shorter (~10 ms) in
Figure 24: Developmental dynamics of song tempo and rhythm (NON-ISO ex.). All data are from bird R5106 (see also: Figs. 22 and 23). a-b, Full developmental trajectories of daily song cycle (a) and B-syllable phase angle (b) distributions, with opacity set at 5% for all data points. In a and b, the NON-ISO model cycle duration and phase, respectively, are plotted as horizontal red reference lines, with the ISO model included as dashed blue lines. c, Daily distributions of the bird’s source and target song cycle durations during the transitional period (shown are day 5 for a baseline, and then every other day between days 9 and 25).

absolute duration than that in the model. However, the bird’s source and target song cycles were similar. Consequently, during this period, the bird was singing two different rhythms (isochronous AAA and non-isochronous ABA – though at a different phase than the NON-ISO model), within nearly the same cycle.
In other words, the bird preserved his original tempo (as copied from the isochronous source song) by committing a rhythm imitation error as he incorporated the target song sequence. Thus even though this bird eventually acquired an accurate replica of the NON-ISO target song, he initially inserted syllable B into the song with an inaccurate rhythm, but one which preserved his own prior rhythmic framework. This isochronous ‘meter’ was sustained for an extended developmental period (Fig. 24c), before the bird gradually shifted to the non-isochronous rhythm of the model (Figs. 24a, b). As with the ISO example above (see Fig. 16d), transitional song cycles were not static but rather shifted from day to day. The daily histograms shown in Fig. 24c indicate here too that the bird’s source and target song cycles were largely overlapping in their distributions during much of the transitional period (and that neither was an accurate copy of either the source or target model tempo).

Our results point to two basic rhythmic operations in song learning: imitation of the model rhythm, and preservation of self-tempo. As the above NON-ISO example shows, self-tempo preservation may sometimes trump imitation accuracy: but for this bird, an early sacrifice of imitation accuracy proved to be a successful strategy, and perhaps maintaining a steady framework actually enabled him to correct his song rhythm later in development. What about the less successful imitators in the NON-ISO group? Perhaps in some cases, birds chose to preserve their own tempo at the expense of copying the model. Can we understand the song development of these birds too in terms of a dynamic coupling between the source and emerging target cycles?

Figure 25 presents developmental data for a NON-ISO bird that achieved good sequence imitation (72% alternating trigrams), but produced his copy of the target song with the same isochronous rhythm of his source song – which he had never heard in the context of the new target song. The developmental data reveal a surprising history behind this outcome. Unlike the previous NON-ISO example, this bird did not start out with an isochronous meter when he first began to incorporate the B-syllable. When ABA trigrams first appeared, B was shorter in duration than A (as in his non-isochronous model), and matched the model’s non-isochronous phase (Fig. 25b), but then gradually lengthened over
Figure 25. Acquisition of isochronous rhythm in a NON-ISO bird. All data are from bird R4844. a, Bird’s source (black) and target (green) song cycle duration over developmental time. b, B-syllable phase development. c, e, Sorted raster plots comparing AAA and ABA sequence timing when the B-syllable first appeared (c, 8-14 days after the model was switched) and after the B-syllable stretched to match the duration of A (e, 20-26 days post-switch). d, f, Distributions of the bird’s source and target song cycles at the two time points.

As the B-syllable lengthened, the phase angle advanced to $\pi$ (Fig. 25b), and the bird’s song cycles stretched to match the duration of his original imitation of the source song (Figs. 25d and 25f). Note that these events also coincided with the overtaking of the target song sequence (ABA and BAB trigrams, Fig. 25a).

A third NON-ISO bird (Fig. 26) improvised a novel song element (‘X’, very short-duration yellow syllable in the raster in Fig. 26c) in addition to B (and therefore never exceeded 25% alternating trigrams); but together, the two syllables preserved the isochronous meter of the source song. Here again we found that the bird’s source and target songs cycles drifted together over time (Fig. 26a). In the
**Figure 26. Acquisition of isochronous rhythm in a NON-ISO bird – with an improvised song element.** All data are from bird R5054. a, Bird’s source (black) and target (green) song cycle duration over developmental time. b, X- and B-syllable phase development (see text). c, Sorted raster plot of mature song bouts (32-25 days post-switch), showing interleaving of AAA and AXBA song sequences. ‘X’ is a novel syllable, colored yellow in the raster plot. d, Distributions of the bird’s source and “target” song cycles.

mature songs, AAA and AXBA sequences were time-locked (Figs. 26c-d). **Figure 26b** shows that the first new syllable to emerge was isochronous in phase, unlike the previous example (Fig. 25b). As an interesting side-note, this occurred directly following two days of almost complete quiet (the gaps in Figs. 26a and 26b are not missing data).

As a final example from the NON-ISO group, **Fig. 27** presents a case where the bird entered several blind alleys during development and ultimately failed to acquire either the sequence or the rhythm of the target model. And yet, his original isochronous rhythm can be identified in his new song. At different time points, extra syllables were added at a variety of different phases (Fig. 27b) causing the “target” song cycle to change erratically (Fig. 27a). The song finally stabilized into a stereotyped sequence of
Figure 27. Phase diversity, false starts, and period-tripling in a NON-ISO bird. All data are from bird R4772. a, Bird’s source (black) and target (green) song cycle duration over developmental time. b, X- and B-syllable phase development. c, d Sorted raster plots comparing AAAA sequence timing at days 20-22 post-switch (c) with AX2 X1BA sequence timing at days 67-69 post switch (d). X1 and X2 are two distinct novel syllables, colored yellow and turquoise in the raster plot. e, Distributions of the early source and late “target” song cycle durations.

AX2 X1B (Fig. 27d). Strikingly, even after many complex developments, this final song was a rhythmic match with the bird’s own AAAA source song in the days preceding the onset of syllable diversification (Fig. 27c). However, rather than fitting into a cycle matching the interval between every other A-syllable, the new song appears to have borrowed a different frame from the source song – the cycle between every fourth A (Fig. 27e).

In sum, despite the apparent diversity in outcomes and developmental trajectories among the NON-ISO birds, we observed a general pattern of self-tempo preservation, as in the (more homogeneous) ISO group. This shared tendency was served by several different rhythm and/or sequence

---

3 We number the syllables in accordance with the order in which they appeared. Before generating syllable X2, this bird went through a transitional period of singing stereotyped sequences of AX1B; see the full development raster in the Appendix.
imitation errors (reminiscent of Liu et al.’s [2004] finding that zebra finches can use multiple strategies to learn the same song): initial non-isochronous phase delay, regression to or emergence of isochrony via syllable elongation or the improvisation of an extra ‘filler’ syllable, and finally period tripling, to accommodate an overabundance of added song elements. While only the first of these learning strategies led to both accurate sequence and rhythm imitation of the target model, each allowed birds to maintain tempo continuity throughout development.

3.4.5 A developmental tradeoff: rhythmic reuse or arrhythmia

In the previous section, we showed evidence that a rhythmic framework plays an important general role in song learning, for the less successful NON-ISO imitators no less than the best NON-ISO imitator or the ISO-trained birds. Recall, however, that three out of nine NON-ISO birds managed to learn the target song as well as the ISO birds (Fig. 12). How do the remaining two successful imitators from the NON-ISO group compare? Recall as well that NON-ISO birds as a group showed arrhythmia in maturity (Fig. 13). However, many NON-ISO birds exhibited no difference in temporal stability between their baseline source and endpoint target songs. Since tempo continuity appears to be a strong developmental constraint, and the (overall-successful) ISO birds had equally low tempo variability at crystallization as at baseline, we hypothesized that tempo stability and tempo continuity might be importantly related. Figure 28 presents developmental trajectories of song cycle median-absolute deviation (M.A.D.) for all birds in both groups.
Figure 28. Song tempo variability development – individual trajectories. Separate panels show the development of source (black) and target (green) song cycle median absolute deviation for each bird (ISO, a; NON-ISO, b). As in Fig. 20, individuals are sorted according to endpoint sequence imitation, top left to bottom right.
On the whole, we see that cycle variability of birds’ target songs (in green) tended to decrease over time. The co-existing source song tempo variability (black) remained relatively low across birds. Moreover, the four NON-ISO examples we considered above all had low target song tempo variability (green) at the end of development (Fig. 28b, top left, middle left, bottom left and center). However, the two second-best imitators in the NON-ISO group did not (Fig. 28b, top center and right). Intriguingly, these birds also produced sudden tempo changes in acquiring the target song (Fig. 29), suggestive of a tradeoff with lasting consequences.

**Figure 29. Lasting arrhythmia in NON-ISO birds with developmental discontinuities in song tempo.** Shown are tempo (top) and tempo variability (bottom) developmental trajectories for three NON-ISO birds that achieved good sequence imitation of the target song model (% alternating trigrams, upper right corner of each panel at top). Source and target cycle median and median absolute deviation (M.A.D.) plotted in black green and, respectively.
3.5 Discussion

We investigated the role of rhythm in imitative vocal learning by experimentally decoupling sequence and rhythm learning during zebra finch song acquisition. Our findings revealed that rhythm is operative in song learning. The success rate of incorporating the new syllable type depended on its duration, relative to that of the previously acquired syllable: imitation was impaired when the new sequence was embedded in a new rhythm. Importantly, the time structure of an external model was often acquired relatively rather than absolutely. Finally, developmental analyses revealed that birds maintained an established cycle as they incorporated new song elements, and often appeared to reuse existing timeslots for new content. Together these findings suggest that zebra finches possess rhythmic frameworks in their songs.

According to prevailing neuronal models in which birdsong learning and production is controlled by premotor sequencing mechanisms (Fee et al., 2004; Jin et al., 2007), the learning task in our experiment should have been equally challenging under both ISO and NON-ISO conditions. The learning impairments and developmental trajectories we observed suggest a mechanism for rhythm-coding in the songbird brain, beyond the putative continuous representation of abstract time within HVC (Hahnloser et al., 2002; Lynch et al., 2016; Picardo et al., 2016) or distributed across song system nuclei (Hamaguchi et al., 2016). In other words, our findings suggest the existence of a rhythm generator in the songbird brain – which either drives the song system circuitry, or perhaps is an emergent property of it.

Neural data will be crucial for shedding light on the physical mechanism of this rhythm generator, and whether song rhythm learning is mediated by local circuit dynamics (i.e., by a discrete central pattern generator) or more macroscopic brain oscillations of synchronized excitability. The cerebellum plays an important role in regulating speech rhythm and fluency (Doupe & Kuhl, 1999), but cerebellar involvement in song learning is currently unknown. The thalamic nucleus Uva is another potential player. Uva forms a thalamocortical loop with both HVC and its afferent NiF (nucleus interface – an important site of sensorimotor integration [Vyssotski et al., 2016]) via feedback connections from midbrain motor and
respiratory areas (Nottebohm et al., 1982; Striedter & Vu, 1998). A recent study (Danish et al., 2017) found periodic, syllable-associated activity during singing in Uva; however, it is unknown how Uva activity tracks with song development. Longitudinal recordings in Uva and/or NiF could help to shed light on the development of a putative rhythm generator. Transient inactivations (Pristine et al., 2016) during critical moments of song development, e.g., when a bird’s target song model is changed, or when the bird’s own song is undergoing changes, might also be a fruitful line of investigation. While our findings seem to suggest the existence of a rhythm template for song learning, we did not address the questions as to when or how such a template is acquired, or how phonology, sequence, and rhythm are orchestrated during development. Some primary auditory neurons in the songbird brain are specially-tuned to temporal features of acoustic stimuli (Nagel & Doupe, 2008), and a recent study (Lampen et al., 2014) found increased ZENK expression in the auditory ‘cortical’ nuclei NCM and CMM (the caudomedial nido- and meso-pallium, respectively) as well as the amygdala in response to “arrhythmic” (time-scrambled) songs, in adult birds. It would be interesting to conduct a similar assay, using our source and target models as stimuli, at different time points in development. Finally, the brainstem, which both controls breathing patterns and represents the only point of interhemispheric connection between the left and right HVC’s (Ashmore et al., 2005), may be instrumental in the development of recurrent information flow in the song system and likewise of rhythmic organization of singing behavior.

We propose that the specific outcomes we observed in our experiments can be explained by simple learning principles, which might have general application (Fig. 30). Our findings suggest two rhythmic operations important in song learning: first, the bird ‘aims’ to imitate the (relative) rhythm of an external model; second, the bird’s aims to maintain an established tempo during development. Tempo is not copied from the model (rhythm is learned relatively rather than absolutely), but the bird’s own tempo imposes a strong constraint on the incorporation of new song elements. The repeating cycle of the bird’s song may also drift in duration over the course of development, but continuity of tempo must be maintained. If sudden tempo discontinuities occur during learning, this leads to lasting arrhythmia in terms of rendition to rendition variability.
Figure 30 provides a phenomenological categorization of the learning outcomes we observed. In our experiment, all birds started out (far left) having already internalized a relative copy of the monosyllabic, isochronous source song, which may be produced at a slower (depicted) or faster tempo. Birds were then switched to either the isochronous or non-isochronous target model which contained an additional syllable (ABAB). As shown, the path to a stabilized song in the ISO condition involves considerably fewer possible trajectories than in the NON-ISO condition. In the most straightforward of trajectories (Figs. 30a and 30c), the bird adopts the new target song concordant with his own source cycle. Notably, the NON-ISO rhythm is not more difficult to acquire than the ISO rhythm in this case (Fig. 30c; see also Figs. 23-24 for NON-ISO data exemplifying this successful outcome). However, in the NON-ISO condition, this outcome comes at the expense of an imitation error, since the NON-ISO source and target song cycles are discordant. If, however, the NON-ISO bird instead prioritizes copying the target model rhythm, the cost is temporal stability (Fig. 30b) – which in itself might be an important outcome of song development, and therefore a high price to pay. In the end, birds may choose to prioritize maintaining concordance between their own source and target song cycles (Fig. 30d-g), thereby achieving temporal precision in maturity, although sacrificing accurate imitation of the target model rhythm (Fig. 30d) and/or sequence (Figs. 30 e-g).

If an ISO bird initially incorporates the B-syllable within a source-discordant cycle (for example if his B-syllable is initially the wrong duration; see Fig. 17), our schema predicts that he will be impelled to rectify this by adjusting his new target cycle, his source cycle, or both, and thereby arrive at a final song that is a good imitation of the relative rhythm of the model. In general, we hypothesize that the rhythmic concordance between the model source and target songs in the ISO condition aids the bird in imitating the model within his preexisting rhythmic framework, while rhythmic discordance between the source model and NON-ISO target model leads to potential conflicts between rhythm copying accuracy and song stability.
Figure 3. Phenomenological categorization of the observed effects of rhythm on song learning. Shown are seven possible trajectories (1 ISO and 6 NON-ISO) toward crystalized song, all following the putative developmental constraint that the bird must maintain his preexisting rhythmic framework. The possession of discordant song cycles is hypothesized to be an unstable situation (b, dashed arrow). The cartoon proceeds from left to right. Except where labeled as the model source or target, all circles represent cycles in the bird’s own song, with circle size indicating cycle duration. Far left, baseline condition in which all birds imitate the AA source song from the model. Concentric circles represent the cycle calculated as one or two repetitions of the A-syllable (for comparison with the two-syllable cycle of the AB target song). In the outer circle, the onset of the second A is indicated by a white bar (at π). The source cycle of the bird is here drawn larger than that of the model as a reminder that the bird may copy the isochronous rhythm of the source song at his own idiosyncratic tempo. Once the source song is acquired, the bird is assigned to either the ISO (blue, upper) or the NON-ISO (red, lower) target condition. See text for explanation of individual trajectories a-g. Deviations from the target model (e-g) are indicated.
A limitation of our experimental design is that the NON-ISO target model deviated from the source in both phase-angle and cycle, whereas the ISO target was concordant with the source model in both respects. This makes it impossible to pinpoint the relative impacts of changing the phase vs. changing the cycle (even though we found that the absolute cycle duration is not necessarily copied). It may be that both contributed to making the sequence learning task easier for the ISO group. In order to test this, we could conduct additional experiments in which the target ABAB model was isochronous but deviated from the source in tempo. If absolute continuity of tempo in the model is necessary for acquiring the new song, then a sped-up (or slowed-down) ISO group should have greater difficulty with the task. Alternatively, if birds are insensitive to tempo changes in the model, then a time-warped ISO group should perform as well as the ISO condition that we used in the present study. Finally, we would also want to test a NON-ISO group in which phase was altered but the source cycle was preserved.

In conclusion, our findings suggest that the developing bird’s current rhythmic framework is the necessary context for incorporating any new song elements. If the model song rhythm is altered, birds may copy rhythmic elements (phases) of the new target, but they appear unable to reproduce the new rhythm in toto without integrating these elements into their existing rhythmic framework. This integration may occur rapidly, by selecting a successful context (cycle) for incorporation; or it may occur piecemeal, through distortion of the model target song rhythm and/or incomplete incorporation of song elements. Regardless, birds can only establish their songs – whatever their fidelity to a model – within a rhythmic framework.
4.1 Abstract

In many songbird species, males sing to attract females and repel rivals. How can gregarious, non-territorial songbirds such as zebra finches, where females have access to numerous males, sustain monogamy? We found that the dopaminergic reward circuitry of zebra finches can simultaneously promote social cohesion and breeding boundaries. Surprisingly, in unmated males but not in females, striatal dopamine neurotransmission was elevated after hearing songs. Behaviorally too, unmated males but not females persistently exchanged mild punishments in return for songs. Song reinforcement diminished when dopamine receptors were blocked. In females, we observed song reinforcement exclusively to the mate’s song, although their striatal dopamine neurotransmission was only slightly elevated. These findings suggest that song-triggered dopaminergic activation serves a dual function in social songbirds: as low-threshold social reinforcement in males and as ultra-selective sexual reinforcement in females. Co-evolution of sexually dimorphic reinforcement systems can explain the coexistence of gregariousness and monogamy.

---

4 A version of this chapter has been published as Tokarev et al., 2017.
4.2 Introduction

Many species of highly gregarious and colonial birds form long-term monogamous pairs (Goodson et al., 2012; Goodson & Kingsbury, 2011; Griffith et al., 2010; Zann, 1994). Pair bonding and flocking behaviors are regulated by neuropeptides and dopaminergic reward system (Goodson et al., 2012; Goodson & Kingsbury, 2011). However, for an animal to be highly social and at the same time monogamous, it must possess two distinct reinforcement systems: one with low selectivity for social stimuli to promote aggregation, and another highly selective for sexual stimuli to promote monogamy. But many communicative stimuli, including birdsong, may serve both social and sexual functions. In such cases, reinforcement may depend on stimulus context: for example, in many solitary songbird males, producing the same song may either attract females or repel rival males (Kroodsma & Byers, 1991; Slater, 2003). In social songbirds, however, many females and males live in close proximity, which gives females immediate access to numerous males whose songs may sexually attract them. What is it, then, that allows gregariousness and monogamy to coexist? We investigated this question in zebra finches, which are highly social, yet monogamous songbirds (Griffith et al., 2010; Zann, 1994). Male zebra finches produce a single stereotyped song that can be female-directed or undirected (Jarvis et al., 1998; Scharff & Nottebohm, 1991; Sossinka & Böhner, 1980; ten Cate, 1985; Woolley & Doupe, 2008). Males typically tolerate the singing behavior of their neighbors even when housed in crowded cages, although the song is occasionally used in an aggressive context too (Ihle et al., 2015). Female zebra finches are attracted to male songs (Holveck & Riebel, 2007), but do not sing (Nottebohm & Arnold, 1976).

The zebra finch striatal dopaminergic reward circuitry is activated in both social and sexual contexts (Banerjee et al., 2013; Ihle et al., 2015; Iwasaki et al., 2014; Sasaki et al., 2006). In general, there are more dopamine-producing neurons in social than in territorial songbirds (Goodson et al., 2009). In zebra finches, gregariousness is correlated with the level of activity in dopaminergic neurons (Kelly & Goodson, 2015). Striatal dopamine increases in social situations, e.g., when adult males interact with females (Ihle et al., 2015; Sasaki et al., 2006), or juvenile males with adult male tutors, and importantly, even without singing in either of these contexts (Ihle et al., 2015). During pair formation striatal dopamine
levels increase in both sexes (Banerjee et al., 2013; Iwasaki et al., 2014). In the context of song learning, striatal dopaminergic input to song control nuclei is modulated during singing (Gadagkar et al., 2016; Hoffmann et al., 2016; Simonyan et al., 2012). However, although song is an important sexual stimulus in songbirds (Kroodsma & Byers, 1991; Slater, 2003), there is no direct evidence that hearing songs may affect striatal dopamine in either sexual or affiliative (Hausberger et al., 1995) context. Here we performed in vivo imaging and behavioral experiments that show the forebrain dopaminergic system response to song stimulation in zebra finches across sexes and breeding states, in order to distinguish between social and sexual components of a putative rewarding (reinforcement) function of song in social songbirds.

We developed two complementary experimental approaches. First, we used a delayed positron emission tomography (PET) procedure (Patel et al., 2008) in order to measure dopamine neurotransmission (Laruelle, 2000) in awake and unrestrained birds. Zebra finches were injected with $[^{11}\text{C}]$raclopride radiotracer, which binds to dopamine type 2 (D2) receptors. Instead of acquiring PET immediately, we first stimulated birds with song playbacks for 20 minutes, while they were awake and freely-behaving, and then scanned them just after the stimulation under general anesthesia (delayed PET, Fig. 31, see protocol in Methods). Second, we developed an apparatus for assessing song reinforcement behaviorally. This approach is a variant on drug addiction experiments, which typically measure how much rodents are willing to work, or exchange mild punishment, in return for access to dopaminergic stimulants such as cocaine (Shaham et al., 2000) (Fig. 32). We used a song stimulus instead of the drug and measured the extent to which birds were willing to receive mildly aversive air puffs (Tokarev & Tchernichovski, 2014) in exchange for hearing song playbacks. Finally, in order to test for causality between dopamine neurotransmission and song reinforcement behavior, we blocked dopamine neurotransmission with a selective antagonist of D2 receptors L-741,626 (Li et al., 2010; Watson et al., 2012). We used PET to determine the localization of dopaminergic blockage, and then tested behaviorally if blocking of dopamine D2 receptors was sufficient to diminish reinforcing effect of songs.
4.3 Method

4.3.1 Experimental design

This study was conducted in accordance with the guidelines of the US National Institutes of Health and was approved by the Institutional Animal Care and Use Committees of Hunter College of the City University of New York and Weill Cornell Medical College.

Eleven adult male and seventeen adult female zebra finches (Taeniopygia guttata) bred at Hunter College (room temperature 19-24˚C, 12:12 h light/dark cycle) were used in the neuroimaging experiments. Animals were raised by both parents until adulthood and were housed, except during these experiments, social groups of 2-20 within the breeding colony room, with possibility to engage in social interactions with other zebra finches. All males and nine of the females were non-mated, eight other females were mated in breeding pairs.

The concept of our work is similar to a human study, where favorite musical pieces were shown to increase striatal dopamine levels (Salimpoor et al., 2011). Here we employed a modification in PET protocol that allowed to obtain measurements that reflected changes in dopamine release in awake songbirds. Before imaging, the non-mated animals were injected $[^{11}\text{C}]{\text{raclopride}}$ and then either exposed to 20 minutes of recorded songs of unfamiliar male zebra finches (one novel song every 15 seconds) or kept in quiet conditions for 20 minutes (Fig. 31). This time interval was chosen according to the $^{11}\text{C}$ half-life of 20 min and its detectability with the current PET technique. PET and anatomical X-ray computed tomography (CT) images were acquired immediately afterwards using an Inveon™ Research Workplace (Siemens). Delayed PET scans for dopamine are well established in several animal species (Marzluff et al., 2012; Patel et al., 2008), but since this is a novel method for measuring striatal responses to birdsong, we describe it in detail as a protocol in the next section.

In the experiments with mated females, stimuli were songs of either their own mate or another mated male; they were also synchronized in their breeding cycle so that during stimulation and PET they
would be in similar hormonal states. The females were kept together with their mates for the first week after hatching of the offspring but then were moved (together with offspring) to the nursery room in the absence of adult males until post-hatch day 30, after which they would reunite with their mates. This cycle is routinely performed in the laboratory to produce juvenile zebra finches not exposed to adult male song, which we use in other studies. For this experiment, we used females that had gone through this cycle several times, and stimulation/scanning took place shortly before their return to the mates. Scanning procedures were the same as in the previous experiment and are described in more detail in the next section.

Eight of the males, four unmated females and four of the mated females were also tested in a behavioral paradigm for preferences to the auditory stimuli that had been used in the PET experiments (Fig. 32, a modification of the socially-reinforced auditory discrimination paradigm described previously [Tokarev & Tchemichovski, 2014]). The day prior to testing, birds were kept in social isolation so that they
**Figure 32. Song reinforcement assay.** An apparatus for testing the amount of aversive air puffs birds were willing to receive in exchange for hearing songs. Birds voluntarily perched next to a window through which they could see a video of a singing bird. Videos were presented either silently (control) or accompanied with song playbacks. When the infrared beam detected the bird perching next to the window, aversive air puffs were delivered in random (unpredictable) intervals (with a likelihood of 12.5% sec).

would be attracted to a video of a male (Ljubičić et al., 2016). The video was played either in silence (20 min) or together with the same auditory stimuli as in the PET experiments: a mix of songs of unfamiliar male zebra finches for the males and unmated females, and songs of unfamiliar males or mates for the mated females (20 min). The order of auditory accompaniment (silence/songs) in each session was random; each animal was tested in 10 sessions. In order to see the video and be closer to source of auditory stimulation, the animals had to sit on a perch that produced air puff in a random manner controlled by Bird Puffer software (http://soundanalysispro.com/bird-puffer). We previously determined that random air puffs with a probability of ~2/minute are well tolerated by the birds. Our software automatically registered the bird’s perching activity, delivered the air puffs, and kept continuous records of air puffs that each bird received. We then analyzed during which stimulation the animals were willing to receive more air puffs.
We also tested whether the movement might account for observed differences in striatal dopamine release. If dopamine level changes were due to movement, then movement should differ across treatments: higher in zebra finch males but not females when hearing songs compared to when they are kept in silence. To test if this were the case, we performed an additional control experiment with a new group of 8 males and 8 females, where we simulated the song vs. silence pre-PET conditions (including transfer to the same room and raclopride injection), and also video tracked birds' movement. We monitored for such body movements as flying, hopping and wing-whirring, as well as quantitatively analyzed the Euclidian distances travelled every 0.3 seconds for the center of body mass and beak to continuously track changes in position of body and head, respectively.

4.3.2 Injections of L-741,626

To detect whether dopamine neurotransmission was necessary for the observed behavioral effects in males, four of them were injected with L-741,626 (Sigma-Aldrich, Saint Louis, MO, USA), a very selective antagonist of D2-receptors, which had been used to study the function of D2-receptors in rodents (Dai et al., 2016; Li et al., 2010; Watson et al., 2012) and primates (Achat-Mendes et al., 2010). We injected L-741,626 intraperitoneally at 3.33µg/g body weight, within the range described for rodents (Li et al., 2010; Watson et al., 2012), diluted in saline (acetic acid was added to increase solubility at first, then pH was neutralized by caustic soda solution). The L-741,626 injections were administered 30 min before each test with at least 48h between treatments, 5 times for each animal, with an intra-individual control of sham injections (saline) of the same volume.

4.3.3 Simultaneous PET on four zebra finches to measure dopamine released during auditory stimulation in awake unrestrained state

We established a minimally invasive method for in vivo imaging in zebra finches to measure dopamine neurotransmission in four awake unrestrained animals simultaneously; these measurements may be taken multiple times allowing for intra-subject comparisons (Fig. 31). Due to their small size compared to the available imaging volume of our micro-PET, we were able to scan four birds simultaneously. Thus, the experiments were done in tetrads, with two animals in one condition, and two
animals in another; the conditions were then reversed in the subsequent PET scan experiment performed. $[^{11}\text{C}]$raclopride was delivered via intravenous (i.v.; ulnar vein) or intraperitoneal (i.p.) bolus injections that lasted around 1 min or less; radioactivity doses were ~300 µCi or less, in solutions of 150 µl for i.p. injections and 100 µl for i.v. injections with $[^{11}\text{C}]$raclopride mass at ~0.3 nmol/g (body weight). Usage of $[^{11}\text{C}]$raclopride to track changes in dopamine levels has been validated in studies with simultaneous microdialysis (Morris et al., 2008; Normandin et al., 2012).

When dopamine is released, decrease in radioactive $[^{11}\text{C}]$raclopride signal is mediated through direct competition between these two molecules for D2 receptors (Fisher et al., 1995) and as a result of D2 receptors switching from low to high affinity for dopamine but not raclopride (Fisher et al., 1995; Seeman et al., 1994); also, the striatal $[^{11}\text{C}]$raclopride signal does not rebound after its decline once dopamine is released (Endres et al., 1997). Therefore, differences in dopamine neurotransmission between zebra finches exposed to song playbacks and silence observed in our work were likely due to experimental conditions, even though imaging was performed after stimulation (Yoder et al., 2008). This method of delayed PET (aka “awake uptake”) was first used to detect changes in dopamine levels in freely moving rats (Patel et al., 2008). A similar protocol was also used in songbirds (crows), but with $[^{18}\text{F}]$-2-fluoro-2-deoxy-D-glucose to detect general brain activation in response to visual stimuli (Marzluff et al., 2012).

The animals were let to recover after handling for 1-2 min and then were kept individually either in quiet conditions (20 min) or were presented with recordings of various zebra finch songs (one novel song every 15 sec during 20 min), thus providing stimulation almost immediately after radioligand injection, similarly to previous studies (Marzluff et al., 2012; Patel et al., 2008). Food and water were provided ad libitum. None of the birds sang or attempted to sing during the 20 min of the experiment (in all conditions). Behavioral activity was at minimum during the experiment with no drinking or feeding observed, and only occasional perching. This suggested that the difference in experimental conditions (song playbacks or silence) would be the sole factor in possible differences in dopamine neurotransmission. Immediately after the experiment, the animals were sedated for ~2 min under 3% isoflurane in O2, 2 L/min, and then
transferred to a custom-made plexiglass chamber with 4 head holders made from vinyl tubes; their bodies were additionally fixed with a surgical tape to reduce spontaneous movements during scanning. Animal placement (2 in radial, 2 in axial direction; heads facing towards the center of the imaging volume) was chosen to maximize image quality (Siepel et al., 2010). The chamber was then placed in the micro-PET scanner, and anesthesia was reduced to 2% isoflurane. Acquisition of the radioactive signal lasted 60 min and was followed by an anatomical CT scan of 10 min duration. Differences in radioactive signal acquired during the PET scan were expected to reflect dopamine release during auditory stimulation, as after $[^{11}C]$raclopride is displaced by dopamine its level does not rebound within this time frame despite clearance of dopamine and even with continuous infusion of $[^{11}C]$raclopride (Endres et al., 1997), while we performed single bolus injection. We were able to inject a sufficient amount of radiotracer to obtain images of $[^{11}C]$raclopride uptake, and all animals recovered quickly after the scan. We established that both i.v. and i.p. injections of $[^{11}C]$raclopride produced a radioactive signal in striatum that was detectable by micro-PET, and the data from birds after i.v. and i.p. injections of $[^{11}C]$raclopride overlapped and therefore were combined. Thus, both injection methods appeared to be effective for detection of dopamine level changes. We recommend i.p. injections for future research, as they are faster and easier to perform, require less handling and thus are less stressful for animals (and experimenters).

We also performed an additional PET scan on four males that had been tested with the D2 receptor antagonist, L-741,626, to confirm that it blocked binding at the receptor. Two birds were injected L-741,626 solution and two others saline 30 min before $[^{11}C]$raclopride injection. The rest of the procedure was the same.

4.3.4 Radiochemistry

The radiotracer $[^{11}C]$raclopride was synthesized on-site immediately before each experiment at the Citigroup Biomedical Imaging Center, Weill Cornell Medical College, following standard procedures (Broft et al., 2015; Mawlawi et al., 2001). The average specific activity of $[^{11}C]$raclopride was 6046 mCi/µmol. $[^{11}C]$raclopride was isolated and formulated into an isotonic solution containing 5-7% ethanol, with concentration of 0.13 µg/mL. Although alcohol could potentially influence behavioral state of the
animal, the amount injected in our experiments (~0.3 g/kg) was substantially lower than that causing an intoxicated stupor in a previous study (2-3 g/kg) (Olson et al., 2014) and importantly was similar across all experimental conditions.

4.3.5 PET image preparation and statistical analysis

PET imaging data were first processed in PMOD software (http://www.pmod.com). As four animals were scanned simultaneously at each experiment, raw images were separated into four zones around each brain and cropped accordingly in PMOD software. PET data were summed across 6 evenly distributed time points for each scan. Further, PET data were processed and analyzed in SPM12 software (http://www.fil.ion.ucl.ac.uk/spm).

Anatomical CT images were transformed into standardized stereotaxic space and aligned with a 3D magnetic resonance imaging atlas of the zebra finch brain, which also references common brain areas (Poirier et al., 2008). All PET images were corrected for volume-to-volume motion by inter-frame realignment and then co-registered to the subject’s anatomical CT image. All alignment transformations were visually inspected to ensure that there was no mismatch with the template brain image. Datasets of three males, one unmated and two mated females were discarded because of difficulties with alignment of the images due to motion during scans. Data from the remaining 22 animals were analyzed further.

[11C]raclopride binding potential for dopamine D2 receptors in each voxel was calculated using a simplified reference region method (Gunn et al., 1997; Lammertsma et al., 1996; Patel et al., 2008), with the cerebellum as the reference region, since it does not contain detectable D2 receptors and is traditionally used for determination of nonspecific binding and free radiotracer in the brain (Lammertsma et al., 1996; Litton et al., 1994): \( \frac{C_{St} - C_{Cb}}{C_{Cb}} \), where \( C_{St} \) is radioactivity concentration in striatal (St) voxels (or anywhere else outside the reference region), and \( C_{Cb} \) is averaged radioactivity concentration in cerebellum (Cb). Therefore, [11C]raclopride binding potential was represented by a striatal-cerebellar ratio (SCR) of radioactive concentrations (Patel et al., 2008). As [11C]raclopride and dopamine compete for D2-receptors, decrease in [11C]raclopride binding potential indicates an increase of dopamine concentration (Endres et al., 1997; Fisher et al., 1995) and thus reflects increased dopamine neurotransmission.
(Laruelle, 2000; Martinez et al., 2003). Statistical parametric maps of [11C]raclopride binding potential change were produced by comparing the parametric SCR maps of the two scan sessions (song playbacks and quiet condition, or mate’s and unfamiliar songs); comparisons between two conditions were performed with paired t-tests, with two-tailed probability value of $p < .05$ chosen as statistically significant (Urban et al., 2012). Clusters of significant change were identified in xjView (http://www.alivelearn.net/xjview) at $p < .05$; p-values corrected for multiple comparisons were calculated for each cluster of contiguous voxels at a t-threshold of 3.56 within a search volume equal to the whole brain and an effective spatial resolution of 1.4 mm full-width at half maximum (FWHM) (Salimpoor et al., 2011). Mean binding potential values were extracted from the significant cluster for each individual, and the normalized percent change in dopamine level was calculated as $\Delta = \frac{(\text{SCR}_{\text{silence}} - \text{SCR}_{\text{song}})}{\text{SCR}_{\text{silence}}} \times 100$.  

4.4 Results

We first tested if our delayed PET technique could detect changes in striatal dopamine neurotransmission after hearing song playbacks. We scanned eight unmated female zebra finches, where we expected to find higher levels of dopamine neurotransmission after song playbacks (i.e., lower levels of $[^{11}C] \text{raclopride}$ binding), and eight unmated males, where we expected to find a weaker effect, if any. Each bird was scanned twice: after stimulation with a variety of unfamiliar songs (both female-directed and undirected) over 20 minutes, and after silence over the same duration (Fig. 31). As expected from the distribution of dopamine receptors in the songbird brain (Kubikova et al., 2010), the averaged PET map showed that the striatum was the major site of $[^{11}C] \text{raclopride}$ binding in both conditions in males (Fig. 33a) and in females (Fig. 33b). However, against our expectations, lower level of $[^{11}C] \text{raclopride}$ binding after hearing songs (suggesting increased striatal dopamine neurotransmission) was detected only in the male group. In males, the song minus silence parametric difference map showed that song stimulation resulted in significantly lower level of $[^{11}C] \text{raclopride}$ binding in a part of the striatum (Fig. 33c; cluster-level $p_{\text{corrected}} = .024$, paired t-test corrected for multiple comparisons). Exploratory analysis of individual changes (within the cluster of significant change) showed that $[^{11}C] \text{raclopride}$ binding was at lower levels in all males after hearing songs by $29 \pm 8\%$ (mean $\pm$ s.e.m. hereafter; Fig. 33d; $p = .015$, pair-wise t-test). These results, based on PET of D2 receptors, are comparable to the $26.5 \pm 8.4\%$ increase in dopamine detected with microdialysis in a study where male zebra finches were presented with females (Ihle et al., 2015), confirming that $[^{11}C] \text{raclopride}$ binding at D2 receptors is a robust indicator of the overall striatal dopamine neurotransmission. Surprisingly, females lacked any brain areas with significant change in $[^{11}C] \text{raclopride}$ binding in response to song playbacks. Nevertheless, we produced a mask image from the cluster of significant change in males (Fig. 33c) and used it as a volume of interest to assess for a possible effect in females. Exploratory analysis of individual changes in females showed no apparent change in striatal $[^{11}C] \text{raclopride}$ binding in response to song playbacks (Fig. 33e; $0.4 \pm 6\%$, $p = .737$, pairwise t-test). A direct comparison between males and females showed statistically significant differences in striatal $[^{11}C] \text{raclopride}$ binding after hearing songs (Fig. 33; $p = .015$, t-test). Note, however,
that the difference in the magnitude of change between males and females is, at least partially, driven by the low baseline (silence) \[^{11}\text{C}]\text{raclopride binding in females (Fig. 33e).}\)
Figure 33. Dopamine neurotransmission in response to song stimuli in unmated males and females. Brain schemas in a-c show: cerebellum (Cb), auditory field L (L2), striatum (St), and song control nuclei Area X (X) and lateral magnocellular nucleus of the anterior nidopallium (LMAN). Section planes are shown as dashed orange lines. a & b, Bright yellow areas represent the Statistical Parametric Map (SPM, intensity threshold at \( t \geq 2 \)) for averaged \([^{11}\text{C}]\text{raclopride binding potential in males (a) and females (b; } N = 8 \text{ in both groups)}. SPM is shown over the brain template magnetic-resonance image. In both males and females \([^{11}\text{C}]\text{raclopride binding was restricted to the striatum). c, SPM of the difference in dopamine neurotransmission as detected by \([^{11}\text{C}]\text{raclopride binding in song and silence conditions in males. SMP reveals significantly lower level of \([^{11}\text{C}]\text{raclopride binding in response to hearing novel conspecific songs in males (pairwise t-statistic, cluster-level } p_{\text{corrected}} = 0.024 \), which indicates higher dopamine neurotransmission in this condition. Significant difference was detected in one cluster within the dorsal striatum, mostly outside Area X. d, Analysis of individual changes in \([^{11}\text{C}]\text{raclopride binding in males, comparing song vs. silence. e, Same for females. As no significant cluster was found in females, males’ cluster was used as a mask to produce individual values of \([^{11}\text{C}]\text{raclopride binding within the same area.}}

The sexually dimorphic striatal response to songs could reflect behavioral or anatomical differences between sexes not related to reinforcement. First, as striatal dopamine neurotransmission correlates with movement (Cousins and Salamone, 1996; Gadagkar et al., 2016; Howe and Dombeck, 2016), we tested if birds tended to move more when hearing song playbacks, in a manner that could explain our results. We analyzed movement in eight males and eight females, in similar conditions to those in our experiments before PET scan: injection of raclopride followed by 20 minutes of silence or song playbacks. We observed very little of such body movements as flying, hopping and wing-whirring, and also quantitatively tracked the whole body movement (analyzed every 0.3 seconds for the center of body mass), but there were no significant differences between conditions or sexes (Table 1). Tracking head movement, we observed a significant trend to move the head more during song playbacks in most birds. However, there was no significant difference between males and females in this respect (Table 2). Therefore, mere movement is unlikely to explain our finding of male-specific dopamine response to songs (Fig. 33).

Another concern is that our results could simply reflect anatomical dimorphism in the basal ganglia pathway of the premotor song system: in particular, Area X, which has high density of dopamine D2 receptors (Kubikova et al., 2010) and receives dopamine during female-directed singing (Sasaki et al., 2006), exists only in zebra finch males. However, Area X was mostly excluded from the cluster of significant change (Fig. 33c), suggesting that its contribution was small, if any. This is in line with the
finding that Area X does not respond to auditory stimulation in awake songbirds, except for error signals during singing (Gadagkar et al., 2016).

Table 1. Results of statistical tests to address the differences in body movement in zebra finch males and females in different conditions: in silence or during conspecific song playbacks. Average Euclidian distance every 0.3 seconds was measured in the videos for the center of body mass. Bold-face numbers indicate significance levels $p \leq .05$.

<table>
<thead>
<tr>
<th>Box's Test of Equality of Covariance Matrices</th>
<th>Box's M</th>
<th>F</th>
<th>df1</th>
<th>df2</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>13.334</td>
<td>3.756</td>
<td>3</td>
<td>35280</td>
<td>.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multivariate Tests (Pillai's Trace)</th>
<th>value</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>body movement</td>
<td>0.175</td>
<td>2.968</td>
<td>.107</td>
</tr>
<tr>
<td>body movement * sex</td>
<td>0.02</td>
<td>0.21</td>
<td>.886</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tests of Between-Subjects Effects</th>
<th>df</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex</td>
<td>1</td>
<td>0.249</td>
<td>.626</td>
</tr>
</tbody>
</table>
Table 2. Results of statistical tests to address the differences in head movement in zebra finch males and females in different conditions: in silence or during conspecific song playbacks. Average Euclidian distance every 0.3 seconds was measured in the videos for the position of the beak. Bold-face numbers indicate significance levels $p \leq 0.05$.

<table>
<thead>
<tr>
<th>Box's Test of Equality of Covariance Matrices</th>
<th>Box's M</th>
<th>F</th>
<th>df1</th>
<th>df2</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.004</td>
<td>1.128</td>
<td>3</td>
<td>35280</td>
<td>.336</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multivariate Tests (Pillai's Trace)</th>
<th>value</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>head movement</td>
<td>0.348</td>
<td>7.468</td>
<td>.016</td>
</tr>
<tr>
<td>head movement $\times$ sex</td>
<td>0.016</td>
<td>0.225</td>
<td>.643</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tests of Between-Subjects Effects</th>
<th>df</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex</td>
<td>1</td>
<td>0.598</td>
<td>.454</td>
</tr>
</tbody>
</table>

Given that the expectation of reward is only one of several scenarios that could explain the unanticipated pattern of striatal dopamine neurotransmission that we observed (Cousins & Salamone, 1996; Gadagkar et al., 2016; Hoffmann et al., 2016; Howe & Dombeck, 2016; Kubikova & Kostal, 2010; Riters, 2011; Salimpoor et al., 2011; Schultz, 2002; Stuber et al., 2008), we developed an independent method for assessing the effect of song reinforcement in male and female zebra finches. In order to directly estimate song reinforcement we paired the song stimulus with a mild punishment. We presented the same birds that had been scanned earlier for dopamine with video of a perching male (Fig. 32). Each bird was presented with two daily sessions of videos over ten days (20 sessions, 20 min each). In ten sessions the video was played in silence, and in the alternating ten sessions, it was accompanied by song playbacks (the same mix of initially unfamiliar songs as in the PET experiments). When a bird perched next to the window facing the video display, it would occasionally receive a mildly aversive air puff, in random intervals and without warning. We assessed reinforcement by measuring the number of...
air puffs the bird was willing to tolerate in return for the stimulus, comparing the silent sessions to the song playback sessions.

We found that males voluntarily received many more air puffs during song playback sessions compared to silent sessions (Fig. 34; \( p = .001 \), paired t-test); they appeared attentive during the sessions but did not show any aggressive behavior. Females, on the other hand, showed little motivation to hear song playbacks: their tendencies to receive air-puffs were moderate and did not differ significantly across song playback and silent sessions (Fig. 34; \( p = .267 \), paired t-test).

![Graph showing air puff rates](image)

**Figure 34. Song reinforcement in unmated males and females.** Rate of air puffs (per hour) birds received during song playback and silent sessions: in males (left) and females (right) \((N = 8\) and \(N = 4\), respectively; \( p \)-values for paired t-test shown).

To test whether the song reinforcement we observed in males was dependent on dopamine neurotransmission, we used the D2 receptor antagonist L-741,626 to interfere with D2 receptors. First, we performed a whole brain PET after injections of L-741,626 in order to determine the localization of dopaminergic blockage: as expected, changes in \([11C]\)raclopride binding were observed exclusively in the striatum (Fig. 35). We found substantially lower levels of the striatal binding of \([11C]\)raclopride after
Figure 35. Blockage of D2 receptor binding activity by L-741,626. Statistical parametric map of average $[^{11}C]$raclopride binding is shown over the zebra finch brain template magnetic-resonance image: after saline injection (top) and L-741,626 injection (bottom) ($N = 2$ in both conditions; t-values on the insert). Sagittal (left), frontal (middle) and transverse (right) sections are shown; dashed light-blue lines show section planes.

L-741,626 injection compared to saline (Fig. 35). Therefore, L-741,626 blocks D2 receptors in the songbird striatum as it does in rodents (Li et al., 2010; Watson et al., 2012) and primates (Achat-Mendes et al., 2010). We then tested song reinforcement in four males with our air-puff apparatus as described before, but after injections of either L-741,626 or saline on alternate sessions. On the days of L-741,626 injections, the animals were still active and approached the video, but stimulation with song playbacks no longer increased the number of air puffs they were willing to receive, while on the alternate days of saline injections, song reinforcement was similar to that of untreated males (Fig. 36; see Table 3 for statistics).
Figure 36. Song reinforcement after dopamine receptor blockage. Number of air puffs (per hour) birds received during silent and song playback sessions: after saline injection (left); after L-741,626 injection (right) ($N = 4$; significant p-values are shown for general linear model for repeated measurements; see Table 3 for statistics).
Table 3. Results of statistical tests to address the differences in tolerance to air puffs in zebra finch males in different conditions: in silence or during conspecific song playbacks after saline injections, or same after injection of dopamine receptor antagonist L-741,626. Bold-face numbers indicate significance levels \( p \leq .05 \).

<table>
<thead>
<tr>
<th>Mauchly’s Test of Sphericity</th>
<th>Mauchly’s W</th>
<th>df</th>
<th>( \chi^2 )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td># air puffs /h</td>
<td>0.022</td>
<td>5</td>
<td>6.604</td>
<td>.318</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tests of Within-Subjects Effects</th>
<th>df</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td># air puffs /h</td>
<td>3</td>
<td>7.96</td>
<td>.007</td>
</tr>
</tbody>
</table>

pair-wise post-hoc LSD tests

- song + saline vs silence + saline: .041
- song + saline vs silence + L-741,626: .015
- song + saline vs song + L-741,626: .023
- silence + saline vs silence + L-741,626: .814
- silence + saline vs song + L-741,626: .394
- song + L-741,626 vs silence + L-741,626: .122

How is it that song stimuli are reinforcing in unmated males but not in unmated females? We hypothesized that the non-selective dopamine neurotransmission by unfamiliar songs in males might reflect a social function, but in females, song reinforcement might be exclusively sexually driven, as a part of the mate choice (Riebel, 2009). A possible explanation to those counterintuitive results is that reinforcement could be much more selective in females. We therefore measured song reinforcement in mated females that were ready to breed (Fig. 37). We compared song reinforcement in three conditions: video accompanied with the songs of their mates, video accompanied with songs of unfamiliar mated
Figure 37. Song reinforcement in mated females. Number of air puffs (per hour) mated females received in exchange for silence, non-mate song (from male mated with another female), and mate’s song (N = 4; significant p-values are shown for general linear model for repeated measurements; see Table 4 for statistics).

males, and video alone. The mated females showed little interest in the videos and minimal motivation to tolerate air puffs in return to hearing non-mate songs. However, they were willing to receive many air puffs in return for hearing their mates’ songs (Fig. 37; see Table 4 for statistics).
Table 4. Results of statistical tests to address the differences in tolerance to air puffs in mated zebra finch females in different conditions: in silence and during playbacks of songs of unfamiliar males or their mates. Bold-face numbers indicate significance levels $p \leq 0.05$.

<table>
<thead>
<tr>
<th>Mauchly’s Test of Sphericity</th>
<th>Mauchly’s W</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td># air puffs /h</td>
<td>0.484</td>
<td>2</td>
<td>1.453</td>
<td>0.484</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tests of Within-Subjects Effects</th>
<th>df</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td># air puffs /h</td>
<td>2</td>
<td>13.139</td>
<td>0.006</td>
</tr>
</tbody>
</table>

**pair-wise post-hoc LSD tests**

<table>
<thead>
<tr>
<th></th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>mate’s song vs silence</td>
<td>.021</td>
</tr>
<tr>
<td>mate’s song vs non-mate song</td>
<td>.049</td>
</tr>
<tr>
<td>non-mate song vs silence</td>
<td>.259</td>
</tr>
</tbody>
</table>

Based on these behavioral results, we tested if the pattern of striatal dopamine neurotransmission would be also mate-selective in these females. Using delayed PET, we compared two sets of stimuli: playbacks of mates’ songs versus playbacks of songs produced by other mated males (in both conditions we played a mix of both female-directed and undirected songs). We detected a cluster of voxels with lower [11C]raclopride binding in response to mate song in a small part of the medial dorsal striatum (Fig. 38 a,b); however, the difference across those voxels did not survive correction for multiple comparisons (Fig. 38b). An exploratory post-hoc analysis of individual differences in the same area found that [11C]raclopride binding was 12 ± 4% lower in response to mate song compared to non-mate song (Fig. 38c; $p = .042$, paired t-test). These differences suggested a weak trend for higher levels of dopamine transmission in response to mates’ songs in females, but this borderline effect should be treated with caution and validated in future studies.
Figure 38. Dopamine neurotransmission in response to song stimuli in mated females. a, Brain schemas as in Fig. 33 a, b. Statistical parametric map (SPM, intensity threshold at $t \geq 2$) for average $[^{11}\text{C}]$raclopride binding is shown over the zebra finch brain template magnetic-resonance image. b, SPM of the difference in $[^{11}\text{C}]$raclopride binding in response to non-mate song and mate’s song in mated females ($N = 6$; pair-wise t statistic, $p < .05$). This cluster, however, did not survive correction for multiple comparisons ($p_{\text{corrected}} = .6$, paired t-test corrected for multiple comparisons). c, Individual changes in $[^{11}\text{C}]$raclopride binding in this insignificant cluster in mated females, non-mate song vs. mate’s song.
4.5 Discussion

We found in the zebra finch an unanticipated pattern of sexual dimorphism in dopaminergic responses to song. In males, stimulation with songs resulted in higher levels of striatal dopamine neurotransmission compared to silence condition. Behaviorally too, unfamiliar song playbacks were strongly reinforcing in males. Blocking striatal dopamine D2 receptors extinguished song reinforcement, suggesting involvement of the striatal dopaminergic reward system. In unmated females, hearing songs did not affect dopamine neurotransmission, and playbacks were not reinforcing behaviorally either. In mated females, mate song was strongly reinforcing, with high specificity, but we observed only slightly higher levels of dopamine neurotransmission in response to mate song compared to non-mate song. Thus, in males, both striatal dopamine neurotransmission and behavioral responses to song playbacks indicate low-threshold and non-specific positive reinforcement. This is consistent with a social, perhaps affiliative function of birdsong to promote aggregation (Hausberger et al., 1995). In females, both behavioral and dopaminergic responses to song were high-threshold and mate-selective, consistent with a sexual function to promote monogamy. However, even though behaviorally mated females showed strong reinforcement to mate song, their striatal dopaminergic responses to mate song were barely detectable. This discrepancy will require further assessment in future studies. Note that there are several open questions about the receptor mechanisms that could account for the sexual dimorphism we observed, including different receptors expression levels, different densities of dopaminergic cells, different reuptake mechanisms and different ratios of D1/D2 receptors. For example, it should be tested whether D1 receptors, which are known to be important for reinforcement (Robbins & Everitt, 1996), are also crucial in the reward mechanism of song in zebra finches.

A simple evolutionary scenario can explain the pattern of sexual dimorphism we observed. Territorial songbird males respond aggressively to intruders and are easy to irritate with conspecific song playbacks (Kroodsma & Byers, 1991; Slater, 2003). Females may show strong preference to certain male song features but are generally attracted to conspecific songs (Kroodsma & Byers, 1991; Slater, 2003). Monogamy could be sustained during an evolutionary transition from the territorial to gregarious behavior.
if males evolved high tolerance to song while females simultaneously co-evolved highly selective reinforcement threshold to songs. Our results are consistent with such a scenario. Future studies could test this hypothesis further by systematic examination of sexual dimorphism across territorial and social species of songbirds, and in species where both sexes sing. We would expect to see a lack of song reinforcement in non-social territorial songbirds, at least outside the breeding period. But possibly, aggressive reactions might also increase brain dopamine. For example, it was shown that fighting cocks (Thompson, 1964) may perceive seeing a potential opponent as a reinforcing stimulus; either they look forward to the fight, or to the anticipation of reward after winning the fight. A similar phenomenon was observed in Siamese fighting fish (Thompson, 1963), specific to dominant males (Baenninger, 1970). As avian species demonstrate a wide range of social structures, the reinforcement value of social cues may vary greatly among them. In sum, a sexually dimorphic activation of the dopaminergic reward circuitry that we observed in our study could provide a joint mechanism for aggregation and pair-bonding, two seemingly conflicting characteristics of the social structure of zebra finches and other gregarious yet monogamous species.
CHAPTER 5
CONCLUSION

Temporal structure is a rich site for investigating the interplay between stereotypy and plasticity in vocal learning and communication. At multiple timescales, we have seen how the rhythms of learned vocalizations both constrain and facilitate structured behavioral variation. We began by describing a potential communicative function in the time structure of zebra finch singing. We then presented experimental evidence for rhythmic organization of developmental song learning, and, finally, identified a potential mechanistic basis for song-mediated social cohesion. We found that at the level of zebra finch song bout structure, rhythmic plasticity is distributed across individuals, while during song development, the juvenile bird filters the influence of an external auditory model through a previously-internalized rhythm template. How do song rhythms become consolidated, both during motif learning and in the emergence of signature bout-structure repertoires? And how do the timescales of song rhythm relate to one another? After all, the natural song-learning environment of the developing bird is the rhythmic yet un-stereotyped time structure of the adult’s song bouts. This dissertation provides a foundation for addressing these questions, as yet unexplored in birdsong research.

Such questions also suggest multiple points of connection with human language that could provide a basis for fruitful comparative research. At the most fundamental level, language is time-dependent (except of course in its written form) because it relies on complex motor coordination. A key basic mechanism of patterned movement in all animals, as Von Holst first showed in the 1930s in behavioral studies of spinal limb coordination, is the interaction of endogenous rhythms of neural excitability, or ‘coupled oscillators’ (Von Holst, 1973; Gallistel, 1980). Oscillatory behavior is a pervasive feature of living systems, and the coupling of oscillators is a likely general mechanism of homeostasis (i.e., dynamic stability), on which life depends (Kelso, 1981). In speech production, interacting endogenous rhythms regulate motor timing: wagging a finger at an instructed rate while speaking leads to spontaneous mutual entrainment between vocalizations and finger movements, and vice versa (Kelso et
Natural speech sounds are universally produced at a rate of 3-8 Hz; a ~5-Hz “speech rhythm” also mediates comprehension and induces cortical entrainment in the perceiver (Peelle & Davis, 2012).

Further, this vocal rhythm may point to evolutionary precursors of speech. One reason often given for studying songbirds is that humans have no close relatives possessing the capacity for vocal learning. Other primates vocalize, but none aside from us modify their own vocalizations based on auditory feedback. However, we are not the only species to communicate with the aid of rhythmic facial expressions or gestures. Suggestively, face-to-face lip-smacking in macaques (for example during allopreening, or in mother-infant interactions), is characterized by the same 3-8 Hz-frequency range as speech production (Ghazanfar & Takashi, 2014). This rhythmic similarity has led to the investigation, through mechanistic and developmental studies (reviewed in Ghazanfar & Takashi, 2014), of the hypothesis that the coupling of voice to orofacial rhythms was an key event in the evolution of language (MacNeilage, 1998).

Convergent lines of evidence indicate that the developmental precursors of speech are also rhythmic in nature. Infants and even fetuses (Minai et al., 2017) are sensitive to prosody: language-specific patterns of intonation, relative timing and stress. According to the “prosodic bootstrapping” hypothesis, these “suprasegmental” acoustic cues may facilitate language acquisition (Nazzi & Ramus, 2003). The same appears to be true of second language acquisition (McDonald, 1997), and adults can use prosodic information to learn an artificial language syntax (Valian & Levitt, 1996). In addition, developmental language impairment is associated with rhythm deficits in perception and production (Gordon et al., 2015). In typical development, infant babbling mimics the rhythm of the native language well before sounds acquire referential meaning (Levitt & Wang, 1991). The onset of babbling also often co-occurs with an efflorescence of rhythmic movements, and in the laboratory infants that had just begun to babble were more likely to accept an offered rattle and shake it rhythmically than younger or older infants (Locke et al., 1995). Interestingly, these newly babbling infants showed a bias for right-handed rattle-shaking, indicating that the left-lateralized motor control necessary for speech is not specific to

---

5 Some songbirds also show embryonic responses to the external acoustic environment (Colombelli-Négrel & Kleindorfer, 2017).
Broca’s area. No wonder then that a rhythmic sensorimotor coupling of hand and mouth links gesture with speech inextricably in mature communication (McNeil, 1992; Iverson & Thelen, 1999).

Rhythmic coordination between individuals is also clearly crucial for communication, both linguistic and non—beginning, of course, within the infant-caregiver dyad. The preverbal infant plays an active role in this: coordinated, bidirectional turn-taking in protoconversations between infants and their caregivers at 4 months predicts 12-month cognitive as well as attachment outcomes (Jaffe et al., 2001). Communication rhythms also furnish a social feedback loop for speech development (Warlaumont et al., 2014). One might go so far as to say that the social context of vocal learning is in essence a rhythmic milieu.

In conclusion, rhythm is intimately involved in human speech and language at the mechanistic, evolutionary, ontogenetic, and interpersonal levels. There are therefore extensive opportunities for expanding the parallels between speech and birdsong (cf. Doupe & Kuhl, 1999) by investigating birdsong’s rhythmic organization, with the potential for affording new insights into human development and communicational disorders.
APPENDIX
SONG DEVELOPMENT RASTER PLOTS

i. R4773 (ISO)
ii. R4720 (ISO)
iii. R5102 (ISO)
iv. R4647 (ISO)
v. R5017 (ISO)
vi. R5186 (ISO)
vii. R5107 (ISO)
viii. R4962 (ISO)
ix. R4960 (ISO)
x. R5106 (NON-ISO)
xi. R4718 (NON-ISO)
xb. R5049 (NON-ISO)
xiii. R4844 (NON-ISO)
xiv. R4961 (NON-ISO)
xv. R5185 (NON-ISO)
xvi. R5054 (NON-ISO)
xvii. R4772 (NON-ISO)
xviii. R5191 (NON-ISO)
i. ISO bird 1
ii. ISO bird 2

R4720 song development

Days post-switch

Song bout time (ms)
iii. ISO bird 3

R5102 song development

Days post-switch

Song bout time (ms)
iv. ISO bird 4

R4647 song development

Days post-switch

Song bout time (ms)
v. ISO bird 5

R5017 song development

Days post-switch

Song bout time (ms)
R5186 song development

Days post-switch

Song bout time (ms)
vii. ISO bird 7
R4962 song development

Days post-switch

Song bout time (ms)
ix. ISO bird 9

R4960 song development

Days post-switch

Song bout time (ms)
x. NON-ISO bird 1

R5106 song development

Days post switch

Song bout time (ms)
xi. NON-ISO bird 2
xii. NON-ISO bird 3

R5049 song development

Days post-switch

Song bout time (ms)
xiii. NON-ISO bird 4
xiv. NON-ISO bird 5
xv. NON-ISO bird 6
R5054 song development

Days post-switch

Song bout time (ms)
xvii. NON-ISO bird 8

R4772 song development

Days post-switch

Song bout time (ms)
xviii. NON-ISO bird 9


